LA PRUEBA DEL NUEVE • LA REESTRUCTURACION DEL GPS



JULIO 2004
5,50 EURO

Edición española de
SCIENTIFIC AMERICAN

EL TIEMPO ANTES DE LA GRAN EXPLOSION

EN TORNO A UNA ECONOMIA DEL HIDROGENO

EL TORIO, COMBUSTIBLE NUCLEAR

VUELVE FREUD

MicroARN





Edición española de
SCIENTIFIC
AMERICAN

Julio de 2004

Número 334













DIRECTOR GENERAL José M.ª Valderas Gallardo
DIRECTORA FINANCIERA Pilar Bronchal Garfella
EDICIONES Juan Pedro Campos Gómez
Laia Torres Casas
PRODUCCIÓN M.ª Cruz Iglesias Capón
Bernat Peso Infante
SECRETARÍA Purificación Mayoral Martínez
ADMINISTRACIÓN Victoria Andrés Laiglesia
SUSCRIPCIONES Concepción Orenes Delgado
Olga Blanco Romero
EDITA Prensa Científica, S. A. Muntaner, 339 pral. 1.ª
08021 Barcelona (España)
Teléfono 934 143 344 Telefax 934 145 413
www.investigacionyciencia.es

SCIENTIFIC AMERICAN

EDITOR IN CHIEF John Rennie
EXECUTIVE EDITOR Mariette DiChristina
MANAGING EDITOR Ricki L. Rusting
NEWS EDITOR Philip M. Yam
SPECIAL PROJECTS EDITOR Gary Stix
SENIOR EDITOR Michelle Press
SENIOR WRITER W. Wayt Gibbs
EDITORS Mark Alpert, Steven Ashley, Graham P. Collins, Steve Mirsky, George Musser y Christine Soares
PRODUCTION EDITOR Richard Hunt
GENERAL MANAGER Michael Florek
VICE PRESIDENT AND MANAGING DIRECTOR, INTERNATIONAL Dean Sanderson
PRESIDENT AND CHIEF EXECUTIVE OFFICER Gretchen G. Teichgraeber
CHAIRMAN John Sargent

DISTRIBUCION

para España:

LOGISTA, S. A.
Aragoneses, 18
(Pol. Ind. Alcobendas)
28108 Alcobendas (Madrid)
Tel. 914 843 900

para los restantes países:

Prensa Científica, S. A.
Muntaner, 339 pral. 1.ª
08021 Barcelona
Teléfono 934 143 344

PUBLICIDAD

GM Publicidad
Edificio Eurobuilding
Juan Ramón Jiménez, 8, 1.ª planta
28036 Madrid
Tel. 912 776 400
Fax 914 097 046

Cataluña:
QUERALTO COMUNICACION
Julián Queraltó
Sant Antoni M.ª Claret, 281 4.º 3.ª
08041 Barcelona
Tel. y fax 933 524 532
Móvil 629 555 703

COLABORADORES DE ESTE NUMERO

Asesoramiento y traducción:

M.ª Rosa Zapatero Osorio: *Procesos de formación en la Vía Láctea actual*; Luis Bou: *La prueba del nueve, La reestructuración del GPS* y *Aventuras problemáticas*; José Manuel García de la Mora: *Vuelve Freud*; Ramón Pascual: *El universo antes de la gran explosión* y *El torio, combustible nuclear*; J. Vilardell: *Hace..., Curiosidades de la física, Apuntes* e *Ideas aplicadas*



Portada: Tom Draper Design
MSX/IPAC/NASA (*fondo*)

SUSCRIPCIONES

Prensa Científica S. A.
Muntaner, 339 pral. 1.ª
08021 Barcelona (España)
Teléfono 934 143 344
Fax 934 145 413

Precios de suscripción:

	Un año	Dos años
España	60,00 euro	110,00 euro
Extranjero	85,00 euro	160,00 euro

Ejemplares sueltos:

Ordinario: 5,50 euro
Extraordinario: 6,00 euro

—El precio de los ejemplares atrasados es el mismo que el de los actuales.



ISSN 0210136X Dep. legal: B. 38.999 – 76

Imprime Rotocayfo-Quebecor, S.A. Ctra. de Caldes, km 3 - 08130 Santa Perpètua de Mogoda (Barcelona)

Printed in Spain - Impreso en España



...cincuenta años

QUÍMICA DE LAS PROTEÍNAS. «En el estudio de la química de las proteínas se plantean dos interrogantes. ¿Cuál es la secuencia de aminoácidos en la cadena peptídica? ¿De qué modo se pliega y repliega esa cadena en el espacio ocupado por la molécula? En este artículo consideramos sólo la segunda cuestión. Para abordar el problema, la técnica más fructífera es la brindada por la difracción de rayos X. —Linus Pauling [y otros]» [*Nota de la Redacción: Pauling recibió el premio Nobel de Química a los pocos meses de la publicación de este artículo.*]

DESERTIZACIÓN. «En 1954 tenemos no una región desertizada por las tormentas de polvo, como el *Dust Bowl* hace veinte años, sino dos. Los terrenos menos propicios de las llanuras suroccidentales, sometidos a la acción del arado durante el auge agrícola de la época de la guerra, ya están en camino de la degradación total. Es fácil culpar de tan angustiosa situación a la sequía, pero la sequía es un rasgo climatológico normal de la Grandes Llanuras del sur. No hay que echar la culpa a los elementos, sino a nuestra renuencia a adaptarnos a ellos. Las tormentas de polvo se han ajustado ceñidamente al patrón de las del decenio de 1930. Durante dos o tres años no se consiguieron cosechas en tierras de fertilidad escasa. En los campos desprotegidos, el suelo fértil expuesto fue arrastrado con cada viento animado de la velocidad suficiente para causar erosión. Esas zonas se ensancharon y se fueron uniendo para formar las dos nuevas extensiones de tierra desertizada por la erosión del viento.»

...cien años

FUTURO. «Los que nos hallamos a comienzos del siglo XX, y más aún la creciente mayoría de los que nacimos tras la publicación de *El origen de las especies*, percibimos que el hombre, y todo su mundo, no es más que el presente estadio de un desarrollo tan grandioso y espléndido que, ante su visión, todas las hazañas de la humanidad menguan hasta las proporciones de castillos en la arena. Al mirar hacia los incontables millones de años a nuestra espalda, comprendemos la firme voluntad de vivir pugnando en el légamo entre marea y marea. Al volvernos de nuevo hacia el futuro, de nuestras mentes se ha desvanecido toda idea de finalidad, toda tendencia al milenarismo. El interrogante de lo que ha de venir tras el hombre es el más fascinante en su persistencia y el más insoluble en cualquier lugar del mundo.

—Herbert G. Wells»



Sir William Ramsay en su laboratorio, 1904.

UN QUÍMICO Y UN ELEMENTO. «A sus cincuenta y dos años de edad, sir William Ramsay, cuyo nombre va asociado al nuevo elemento radio, es uno de los más jóvenes científicos del mundo. Puede decirse que su notoriedad pública empezó con sus brillantes descubrimientos de los componentes desconocidos, e insospechados, de la atmósfera (argón, helio, neón, criptón y xenón), descubrimientos efectuados en parte con la colaboración de Lord Rayleigh. La fotografía de Sir William Ramsay fue tomada en su laboratorio especialmente para *Scientific American.*» [*Nota de la Redacción: Ramsay recibió el premio Nobel de Química a los pocos meses de la publicación de este artículo.*]

ESTABILIZADOR DE BUQUES. «El cabeceo de un barco en un mar agitado es ciertamente un grave inconveniente para el bienestar físico de pasajeros y tripulación y para la ejecución de las tareas de a bordo. Ahora, Otto Schlick, conocido ingeniero naval de Hamburgo (Alemania), ha presentado un ingenioso mecanismo ideado para reducir la amplitud de la oscilación. Se basa el aparato en el efecto giroscópico de un volante de inercia, que se monta a bordo del vapor y gira merced a un motor eléctrico.»

...ciento cincuenta años

¿UNA INVENCIÓN INÚTIL? «Escribe el corresponsal en París del ‘New York Times’: ‘Un inventor, que ya se veía acariciando el éxito, acaba de caer víctima de su propia máquina. Esta era un vehículo de vapor que recorría los caminos de posta de Francia. Al descender de una loma, el vehículo del señor Leroy chocó con un obstáculo, volcó y vertió sobre él el contenido de la caldera. Se abrasó. Había invertido diez años y todo su dinero en perfeccionar el invento.’ Fue realmente insensato despilfarrar su dinero en semejante ingenio. Crear ahora carruajes de vapor para los caminos ordinarios, con los ferrocarriles y las locomotoras ya inventados, es como ir a moler con un saco con una piedra en un extremo para equilibrar el saco con el grano en el otro.»

APUNTES

DIABETES

De beta a beta

En la diabetes de tipo 1, la que se declara en la juventud, el sistema inmunitario destruye las células beta, productoras de insulina, y causa una dependencia de por vida de la terapia insulínica. Parecía posible reemplazar esas células y, por tanto, curar la diabetes: estudios anteriores daban a entender que el cuerpo posee células madre en el páncreas que las originan. Unos investigadores de la Universidad de Harvard apuntan ahora que las células beta se duplican por sí mismas. Prepararon ratones para que sus células beta estuviesen dotadas de un marcador genético que podían activar mediante el fármaco tamoxifen. Al administrárselo a los ratones, todas las células beta nuevas poseían genes con el marcador activado; significaba que procedían de células beta preexistentes. Cabe todavía que haya células madre pancreáticas, pero quizá generen sólo una fracción pequeña de las células beta. Si las células beta humanas se originan como las de los ratones, podría conseguirse un tratamiento de la diabetes reforzando el crecimiento de las células beta que sobrevivan.

—Charles Choi



Podría prescindirse de la insulina inyectada si fuera posible reponerse las células beta.

EXTINCIONES

Percusión pérmica

Puede que las huellas del impacto que quizá desencadenó la mayor extinción de la historia de la Tierra se encuentren en Australia. La extinción masiva de finales del Pérmico, hace 250 millones de años, aniquiló siete de cada diez especies terrestres y nueve de cada diez marinas; fue mucho peor que la extinción del paso del Cretáceo al Terciario que extinguió los dinosaurios hace 65 millones de años. Parece que un cráter enterrado de unos 200 kilómetros de ancho, a la altura de la costa noroccidental australiana, es el resultado de la colisión con un meteorito del tamaño del monte Everest y no del volcanismo como hasta ahora se creía. Especímenes de roca del suelo marino que se remontan a finales del Pérmico contienen vidrio en el seno de cristal, característica que se debería a la fusión inducida por las ondas de choque generadas en el impacto. Se ha descubierto también cuarzo fracturado en múltiples direcciones, posible indicio de una colisión cósmica, pues la actividad volcánica fractura el cuarzo en una única dirección. Los supuestos impactos pérmico y cretáceo-terciario podrían ambos haber iniciado un volcanismo a gran escala.

—Charles Choi



El lugar del impacto del meteorito se muestra en rojo, entre los continentes, tal como éstos se disponían hace 250 millones de años.

ARQUEOLOGIA

Cival

En Cival, en la jungla de Guatemala, se va descubriendo poco a poco una ciudad maya. Francisco Estrada-Belli, de la Universidad Vanderbilt, tras una campaña preliminar en 2002, acaba de sacar a la luz dos máscaras de estuco, de 4,5 metros por 3, talladas seguramente hará cerca de 2200 años. Una representa un rostro con la boca adornada de dientes de serpiente. Ha hallado también más de 120 objetos de jade pulido e inscripciones grabadas en una placa de piedra. La ciudad de Cival estuvo habitada desde el siglo sexto antes de nuestra era. Debieron de vivir allí quizá diez mil personas. Siete siglos después la abandonarían. La importancia de estos descubrimientos estriba en que testimonian el esplendor de la civilización maya antes incluso de su época clásica, que empezó en el siglo III después de Cristo.



Máscaras halladas en una ciudad abandonada hace casi dos mil años.

CONEYL JAY *Science Photo Library (arriba)*; LUCY READING, A PARTIR DE UN MAPA DE CHRISTOPHER R. SCOTESE *(centro)*

MEDIO AMBIENTE

Mutaciones del aire



La contaminación del aire afecta a los genes.

La contaminación atmosférica provoca cambios hereditarios, según estudios efectuados en aves y roedores. Para averiguar qué componentes del aire contaminado son mutágenos, se han expuesto dos grupos de ratones de laboratorio al aire de un lugar cercano a dos acerías y una gran autopista. Uno de los grupos, sin embargo, respiró aire pasado por un filtro HEPA (un filtro de las partículas del aire). El experimento se repitió en una zona rural con otros dos grupos de ratones. A las diez semanas de exposición se hizo que criaran. Las crías de los ratones que habían respirado aire contaminado sin filtrar heredaron de sus progenitores mutaciones con una frecuencia doble que las crías de cualquiera de los otros tres grupos. La culpa podría ser de las partículas microscópicas de carbonilla y polvo suspendidas en el aire; suelen llevar adheridas hidrocarburos aromáticos policíclicos tóxicos.

—Charles Choi

CLIMATOLOGIA

Se solventa una polémica

Uno de los debates relativos al calentamiento global se refiere a discrepancias entre datos. Las lecturas de los satélites en la troposfera (la capa atmosférica más cercana a la Tierra) mostraban una tendencia al calentamiento menor que un 0,1 grados Celsius por década, muy inferior al indicado por las temperaturas tomadas en tierra. Los datos de los satélites necesitaban una corrección: la estratosfera, situada sobre la troposfera, disfraza la tendencia al calentamiento. Un grupo de la Universidad de Washington ha analizado las mediciones de 1979 a 2001 de los satélites de la Administración Nacional de EE.UU. para la Atmósfera y los Océanos. Esas sondas medían las microondas emitidas por el oxígeno atmosférico para determinar su temperatura. Aproximadamente un quinto de las señales de la troposfera procedía en realidad de la estratosfera, que se enfría unas cinco veces más rápido que se calienta la troposfera. Compensado ese efecto estratosférico, las lecturas de los satélites se asemejan mucho a las mediciones de las temperaturas en tierra: ambas establecen un calentamiento global general de unos 0,17 grados Celsius por década.

—Charles Choi

COSMOLOGIA

e^{60}

Los antípodas del universo visible distan entre sí actualmente cerca de cien mil millones de años luz. La radiación de microondas que nos llega desde esos puntos diametralmente opuestos se emitió cuando estaban unas mil veces más próximos entre sí que ahora. El universo tenía entonces poco menos de cuatrocientos mil años. No había habido tiempo, pues, para que interaccionasen entre sí. Sin embargo, hoy tienen la misma temperatura. El cielo entero de las microondas cósmicas presenta la misma temperatura, salvo fluctuaciones de menor orden de magnitud. Ese equilibrio térmico da a entender que en algún momento esas partes opuestas del cielo debieron de estar mucho más cerca de lo que la extrapolación de la expansión del universo hasta las inmediaciones de la gran explosión indica. Tuvieron que estar, en concreto, comparando con esa extrapolación, alrededor de 10^{26} veces más cerca. La solución que se ha encontrado es suponer que, muy temprano, el universo se expandió durante un breve período a un ritmo exponencial, es decir, como el número *e* elevado a la constante de Hubble de entonces —el parámetro que mide la velocidad a la que se expande el universo— y el tiempo. Por eso resulta natural calibrar el crecimiento exponencial del que hoy es nuestro universo visible mediante el número de veces que su tamaño se multiplicó por *e*. 10^{26} es igual a e^{60}. Pero, en principio, aquella fase de inflación podría haber multiplicado por *e* un número de veces mucho mayor que 60, cifra que, al fin y al cabo, es sólo una cota inferior dictada por el tamaño del universo visible en un momento de la historia cósmica. Sin embargo, se está argumentando que, si la expansión del universo sigue acelerándose, como está haciendo ahora, en la fase de inflación no pudo haber más que unas 60 multiplicaciones por *e*, o, al menos, que nunca podrá observarse más que el efecto de las últimas sesenta y tantas multiplicaciones por *e* en el fondo de microondas, cuyas fluctuaciones son de una escala tanto mayor cuanto más antiguos fueran sus precursores cuánticos y mayor fuera, por tanto, el número de veces *e* que se amplificaran durante la inflación. Además, dentro de un billón de años se habrían borrado incluso las fluctuaciones visibles hoy. De ser eso cierto, el mismo número que explica la actual homogeneidad general de los cielos marca que lo que hoy vemos en la radiación de fondo es todo lo que se podrá ver alguna vez. ¿Qué tendría de especial nuestro presente, resumido en un 60?

ROBERT LANDAU *Corbis* (*arriba*)

Procesos de formación en la Vía Láctea actual

La Vía Láctea, a la que se suponía una reliquia de un pasado lejano, constituye un objeto vivo y dinámico

Bart P. Wakker y Philipp Richter

Cuando observamos otras galaxias, apreciamos su aspecto global, pero no el detalle. Lo contrario ocurre cuando miramos la nuestra. Inmersos en ella, su estructura general sólo se nos manifiesta indirectamente.

Por eso hemos avanzado muy despacio en el estudio de su configuración e historia. No se tuvo completa seguridad de que fuera un objeto celeste más, uno entre muchos miles de millones, hasta el decenio de 1920. Para mediados de los años cincuenta, ya contábamos, tras muchos esfuerzos, con la imagen que hoy le atribuimos: la de una majestuosa espiral de estrellas y gas. En la década siguiente, se concluyó que se formó en una fase temprana de la historia del cosmos —la estimación más reciente le atribuye 13.000 millones de años— y ha permanecido sin cambios importantes desde entonces.

Sin embargo, poco a poco va quedando patente que la Vía Láctea no es una obra acabada, sino un cuerpo que sigue construyéndose. Como ocurriera con descubrimientos anteriores, lo sabemos en gran medida gracias a la observación de otras galaxias y a la aplicación de lo así aprendido a la nuestra. Se piensa que la mayoría de las galaxias han nacido de la mezcla de precursoras menores. La Vía Láctea exhibe hoy los pasos finales de este proceso: está desmantelando dos galaxias satélites de pequeño tamaño y se apodera de sus estrellas. Mientras tanto, le llegan sin cesar nubes de gas procedentes del espacio intergaláctico. No cabe ya hablar de la formación galáctica en pasado.



LA VIA LACTEA ENGULLE GAS y devora a sus vecinas de menor tamaño. Sigue, pues, formándose. Para entender esta imagen, véase el recuadro "Nuestra galaxia y sus alrededores".

RON MILLER

Las nubes de alta velocidad constituyen la prueba de que la Vía Láctea se agrega gas de manera continua. Estas misteriosas acumulaciones de hidrógeno de hasta 10 millones de veces la masa del Sol y un diámetro de 10.000 años-luz se mueven deprisa en las inmediaciones de la galaxia. Se las descubrió hace 41 años; pero sólo en los últimos cinco, nuevos datos e ideas han enseñado que consisten en gas que se precipita sobre la galaxia. Las nubes demuestran también que la Vía Láctea respira: expulsa gas para atraerlo de nuevo, como si espirara e inspirara. Además, sus propiedades apuntan a que una esfera gigante de plasma tenue y caliente circunda la galaxia. Se sospechaba la existencia de esta esfera, aunque muy pocos pensaban que pudiera ser tan grande.

Ha costado interpretar las nubes de alta velocidad. Como estamos sumergidos en la galaxia, no tenemos una forma directa de saber dónde se encuentran. Podemos apreciar su posición bidimensional en el cielo, pero carecemos de una visión de profundidad. Durante los últimos cuarenta años, esta ambigüedad ha propiciado una diversidad de hipótesis; unas situaban las nubes de alta velocidad cerca del sistema solar, otras en el espacio intergaláctico profundo. Les debemos los descubrimientos más recientes a los telescopios terrestres y espaciales, que han conseguido por fin la localización tridimensional de las nubes.

¿Original o reciclado?

Nuestra galaxia contiene alrededor de 100.000 millones de estrellas, la mayoría concentrada en un disco fino de unos 100.000 años-luz de diámetro y unos 3.000 años-luz de espesor. Describen alrededor del centro galáctico órbitas casi circulares. El Sol gira a una velocidad de casi 200 kilómetros por segundo. Otros 10.000 millones de estrellas forman el esférico "halo" galáctico en torno al disco. Entre las estrellas hay gas y polvo; la mayor parte de este medio interestelar traza también órbitas circulares alrededor del centro galáctico y se concentra en un disco más fino aún que el de las estrellas. Al igual que la atmósfera de un planeta, el gas del medio interestelar es más denso "en el fondo" (el plano galáctico) y se enrarece con la altura. Pero hasta un 10 por ciento del medio interestelar se encuentra fuera del plano y se desplaza unos 400 kilómetros por segundo más rápido de lo que le correspondería por la pura rotación. Ese es el gas de las nubes de alta velocidad.

Empezamos a saber de las nubes de alta velocidad a mediados de los años cincuenta, cuando Guido Münch, del Instituto Tecnológico de California (Caltech), descubrió unas bolsas densas de gas fuera del plano, una clara excepción a la regla de que la densidad del gas disminuye con la altura. Abandonadas a su propia suerte, habrían debido disiparse rápidamente. En 1956, Lyman Spitzer, Jr., de la Universidad de Princeton, presentó una explicación: las estabilizaba una corona gaseosa y caliente que rodearía a la Vía Láctea, una versión a escala galáctica de la corona solar [*véase* "Las coronas de las galaxias", por Klaas S. de Boer y Blair D. Savage; INVESTIGACIÓN Y CIENCIA, octubre de 1982].

Inspirado por la propuesta de Spitzer, Jan Oort, de la Universidad holandesa de Leiden, conjeturó que el halo contenía también gas frío muy lejos del plano de la galaxia. Gracias a sus radioemisiones, se dio con esas nubes frías en 1963. Al contrario que el gas encontrado por Münch, no siguen la rotación global de la galaxia. Parecen estar cayendo hacia el disco galáctico a gran velocidad; se las llamó por eso "nubes de alta velocidad". Ese mismo año se descubrió también un tipo de nube que, aunque se mueve más despacio, no deja de ser anómalo: las "nubes de velocidad intermedia".

Oort elaboró más tarde su idea: constituida ya la Vía Láctea, debió de sobrar gas cerca del borde de la esfera donde la atracción gravitatoria de aquélla se deja sentir; ese gas alcanzaría el disco pasados unos 10.000 millones de años o más, y se presentaría como nube de alta velocidad. Esta hipótesis se ajusta bastante bien a los modelos que explican la composición química de la galaxia. Las estrellas producen elementos pesados y, al morir, los expulsan al medio interestelar. Las estrellas nuevas se los incorporan y producen aún más. Por tanto, si la galaxia evolucionara de manera aislada, cada generación de estrellas debería contener más elementos pesados que las generaciones predecesoras.

Sin embargo, la mayoría de las estrellas de la vecindad solar, sea cual sea su edad, tienen casi las mismas proporciones de elementos pesados. La razón más probable: que la galaxia no está sola y el gas interestelar se renueva sin cesar con gas sin enriquecer. Se sospechaba que algunas o todas las nubes de alta velocidad aportaban ese gas "limpio", pero se carecía de una observación directa que lo confirmase.

Según otra hipótesis, las nubes de alta velocidad no tienen nada que

Resumen/Nubes de alta velocidad

- Desde principios del decenio de 1960, se pensaba que la Vía Láctea y las demás galaxias se habían formado en una fase temprana de la historia del cosmos y habían sufrido luego una lenta evolución. Hoy día, sin embargo, se cuenta con pruebas de que continúan creciendo. Devoran a sus hermanas menores y absorben gas límpido del espacio intergaláctico.
- En nuestra Vía Láctea, asistimos de cerca a ese proceso. El gas que entra adopta la forma de nubes de alta velocidad. Fueron descubiertas hace unas décadas. En fecha reciente, se ha demostrado que algunas llevan materia primigenia; en las observaciones se confunden con el gas circulante.
- Esas nubes tienen apariencias variadas: entre ellas hay acumulaciones de hidrógeno neutro que recuerdan al gas intergaláctico, una corriente de gas arrancada de las pequeñas galaxias cercanas, y gas muy caliente y ionizado que quizá esté disperso por el entorno intergaláctico.

NUESTRA GALAXIA Y SUS ALREDEDORES

CORONA GALACTICA
Gas caliente que rodea a la Vía Láctea

NUBE DE ALTA VELOCIDAD
Acumulación de gas de hidrógeno limpio que cae hacia la galaxia

DISCO GALACTICO
Sistema plano de estrellas, gas y polvo

SUPERBURBUJA
Gas expulsado por las supernovas, brazo anterior del surtidor

NUBE DE VELOCIDAD INTERMEDIA
Gas recondensado, brazo posterior de la fuente

GRAN NUBE DE MAGALLANES
Galaxia satélite de la Vía Láctea

PEQUEÑA NUBE DE MAGALLANES
Galaxia satélite de la Vía Láctea

GALAXIA ENANA ESFEROIDAL SAGITARIO
Galaxia satélite de la Vía Láctea

GALAXIA ANDROMEDA
La mayor galaxia espiral cercana

SOL Y PLANETAS

GALAXIA TRIANGULO
Galaxia espiral cercana de tamaño medio

LA CORRIENTE DE MAGALLANES
Filamento gaseoso arrancado de las Nubes de Magallanes

LA CORRIENTE DE SAGITARIO
Flujo de estrellas arrancadas de la galaxia Sagitario

SLIM FILMS

ver con el flujo de gas hacia el interior, sino que serían parte de un "surtidor galáctico". Esta idea fue propuesta a mediados de los años setenta por Paul Shapiro, ahora en la Universidad de Texas en Austin, y por George B. Field, del Centro Smithsoniano de Astrofísica de Harvard. El gas calentado y ionizado por las estrellas de gran masa asciende desde el disco hacia la corona y crea una especie de atmósfera. Algunas regiones se enfrían y regresan de nuevo, eléctricamente neutras, al disco. Se cerraría así un ciclo de gas entre el disco y la corona. En 1980, Joel Bregman, ahora en la Universidad de Michigan en Ann Arbor, sugirió que las nubes de alta velocidad son ese gas que regresa. Durante algún tiempo, fue la idea más aceptada.

Mareas galácticas

No obstante, ni la hipótesis de Oort ni el modelo del surtidor explicaban todas las propiedades de las nubes de alta velocidad. La situación se complicó aún más a principio de los años setenta con el descubrimiento de la corriente de Magallanes, un filamento de gas que se extiende alrededor de la galaxia. Sigue la órbita de las Nubes de Magallanes, dos pequeñas galaxias compañeras de la nuestra, a la que rodean como las lunas a su planeta. Aunque la palabra "nube" suele referirse a los conglomerados de gas y polvo, estas galaxias repletas de miles de millones de estrellas se llaman así porque en el cielo nocturno aparecen como nubes. Están situadas a unos 150.000 años-luz de nuestra galaxia. En sus muy oblongas trayectorias, ésa es casi la mínima distancia a que pueden hallarse.

La corriente se comporta en muchos aspectos como una sarta de nubes de alta velocidad. Gran parte de ella se desplaza a velocidades incompatibles con una rotación galáctica normal. No cabe explicarlo con ninguna de las dos hipótesis anteriores. De acuerdo con los modelos más detallados de su estructura, los publicados en 1996 por Lance T. Gardiner, de la Universidad Sun Moon de Corea del Sur, y por

Masafumi Noguchi, de la Universidad Tohoku de Japón, el filamento equivale en nuestra galaxia a las corrientes de marea que se han observado alrededor de muchas otras galaxias. Cuando las Nubes de Magallanes se aproximaron por última vez tanto a la Vía Láctea, hace unos 2200 millones de años, la fuerza combinada de nuestra galaxia y de la Gran Nube de Magallanes arrancó parte del gas de las regiones exteriores de la Pequeña Nube de Magallanes. Más o menos la mitad del gas se desaceleró y se rezagó con respecto a las Nubes de Magallanes, en su misma órbita. La otra mitad se aceleró, adelantó a las galaxias y formó el "brazo anterior". Un proceso similar podría estar desgarrando las demás galaxias satélites de la Vía Láctea (*véase el recuadro* "Corrientes").

Otro modelo atribuye la generación de la corriente a las fuerzas de fricción. Si la corona de la Vía Láctea fuese muy extensa, mucho más de lo que supuso Spitzer, le arrancaría gas a las Nubes de Magallanes. En cualquier caso, ambos modelos dictaminan que las Nubes de Magallanes han perdido una gran cantidad de gas; con él se habrían formado muchas de las nubes de alta velocidad.

En 1999 se produjo otro giro importante en la historia de las nubes de alta velocidad. Leo Blitz, de la Universidad de California en Berkeley, y sus colaboradores plantearon que podrían estar mucho más lejos de lo que se había estado creyendo. No surcarían los alrededores de la Vía Láctea, sino las soledades intergalácticas del Grupo Local de galaxias, constituido por la Vía Láctea, Andrómeda y otras 40 galaxias de menor tamaño. El Grupo tiene un diámetro de casi 4 millones de años-luz. En tal caso, las nubes serían restos de la formación del Grupo Local, no de la formación de nuestra galaxia.

Se habían ido proponiendo ideas similares a lo largo de estos últimos 30 años y se las había ido descartando porque las nubes de gas, a tales distancias, carecerían de estabilidad. Pero Blitz conjeturó que las nubes de alta velocidad no son, en realidad, nubes, sino agrupaciones de materia oscura mezclada con una pequeña cantidad de gas. Entonces pesarían diez veces más de lo que se había supuesto y podrían mantenerse unidas por sí mismas. Esta hipótesis despejaría uno de los grandes misterios de la astronomía: los modelos de formación de galaxias predicen bastante más materia oscura residual de la que se ha encontrado [*véase* "El ciclo vital de las galaxias", por Guinevere Kauffmann y Frank van den Bosch; INVESTIGACIÓN Y CIENCIA, agosto de 2002]. Esa materia oscura que falta se hallaría en las nubes de alta velocidad.

Gas caliente

Por tanto, el nuevo milenio empezó con cuatro hipótesis para explicar las nubes de alta velocidad: la del gas primigenio sobrante de la formación de las galaxias, la del gas que circula en un surtidor galáctico, la de los despojos de las Nubes de Magallanes y la de la amalgama intergaláctica de materia oscura y gas. Cada una disponía de algunas pruebas a su favor; se necesitaban datos nuevos para salir del punto muerto. En eso, se ha progresado mucho desde mediados del decenio pasado.

Primero, se completó el mapa del cielo tal y como se lo observa en la emisión de radio del hidrógeno neutro, que revela la distribución del gas con temperaturas en torno a los 100 kelvin. Aad Hulsbosch, de la Universidad de Nimega, y uno de nosotros (Wakker), con el radiotelescopio de Dwingeloo, en Holanda, terminamos la mitad norte de esta exploración en 1988. Ricardo Morras y sus colaboradores cubrieron el cielo sur en 2000 con el radiotelescopio argentino de Villa Elisa (*véase el recuadro* "El cielo de las nubes"). En 1997, Dap Hartmann y Butler Burton, del Observatorio de Leiden, cartografiaron todo el hidrógeno neutro de la Vía Láctea, incluidas las nubes de alta velocidad y velocidad intermedia.

Una contribución más vino por parte de las observaciones en luz visible del "Cartógrafo Wisconsin del Hidrógeno Alfa" [*véase* "El gas entre las estrellas", por Ronald J. Reynolds; INVESTIGACIÓN Y CIENCIA, marzo de 2002]. El hidrógeno neutro no despide luz con longitudes de onda de la zona visible del espectro. El gas ionizado, sí. Las capas exteriores de las nubes de alta velocidad están ionizadas por la luz ultravioleta procedente de la Vía Láctea y otros objetos celestes; las calienta hasta los 8000 kelvin. La cantidad de luz visible es una medida de la intensidad del campo de radiación que rodea a la nube de alta velocidad, que a su vez depende de la distancia al disco galáctico. Por tanto, estas observaciones ofrecen una determinación aproximada de dónde se encuentran las nubes de alta velocidad.

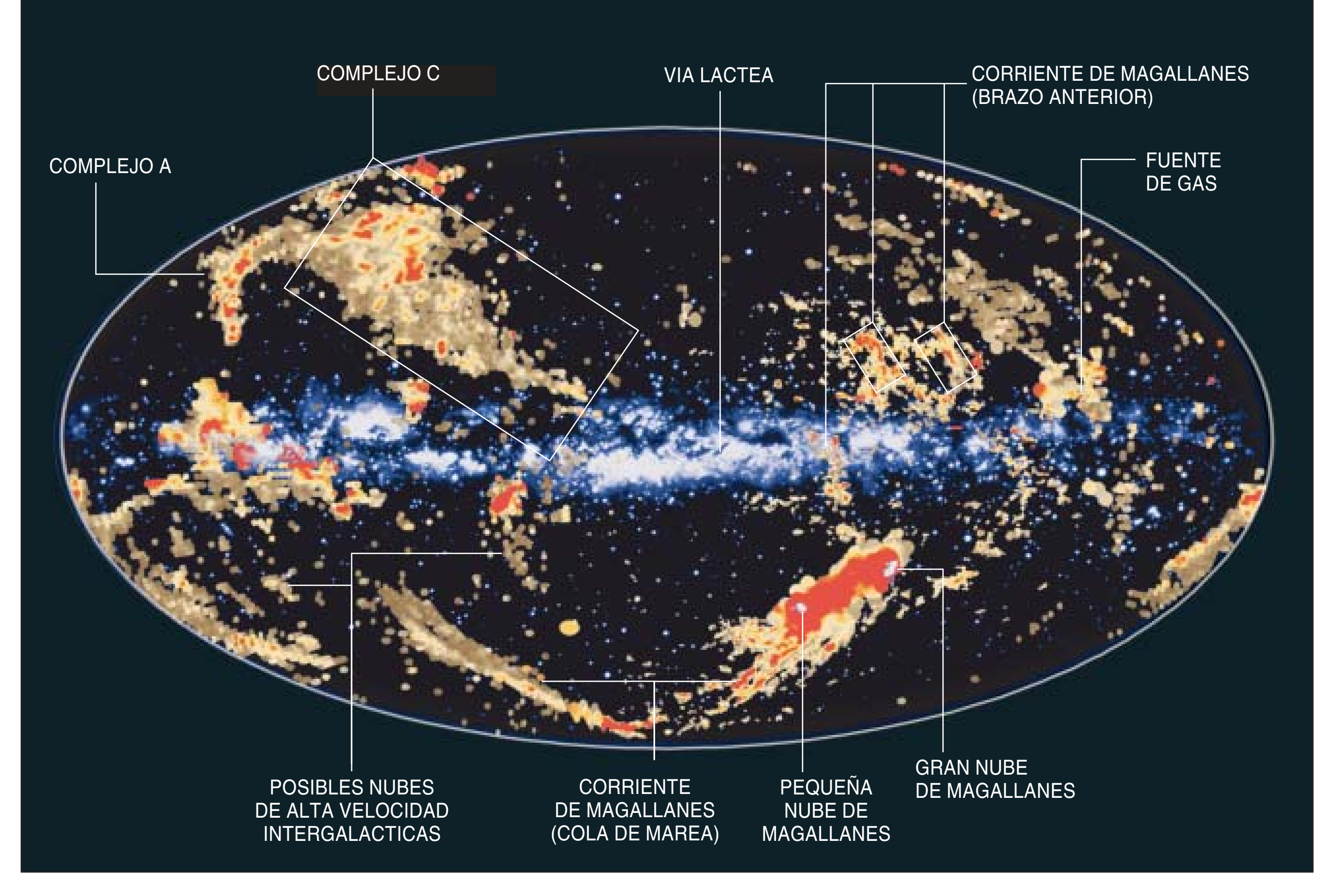
El progreso más espectacular se ha debido a la observación de las líneas de absorción espectrales de las nubes. En lugar de analizar la luz emitida por el gas, estudia la que bloquea (los átomos filtran ciertas longitudes de onda). Los tres observatorios que más han contribuido ahí son: el observatorio del Roque de los Muchachos, en la isla de La Palma, el Telescopio Espacial Hubble y el Explorador Espectroscópico del Ultravioleta Lejano (FUSE son sus siglas en inglés), lanzado en 1999.

Con estos datos, Laura Danly, ahora en la Universidad de Denver, y sus colaboradores pusieron hace 11 años límites a la distancia de una nube de velocidad intermedia. Más recientemente, el equipo de Hugo van Woerden, de la Universidad de Groninga, en Holanda, han medido la distancia a una nube de alta velocidad (*véase el recuadro* "Tras las nubes"). Mientras tanto, nosotros establecimos la composición química de las nubes, completando la información necesaria para discriminar entre las cuatro hipótesis.

Los datos de FUSE muestran una componente muy caliente de las nubes de alta velocidad. Detectó absorción debida a oxígeno altamente ionizado; en concreto, a átomos de oxígeno que han perdido cinco de sus ocho electrones, lo que implica temperaturas del orden de 300.000 kelvin. Estas temperaturas se alcanzan cuando el hidrógeno neutro frío (100 kelvin) entra en contacto con gas muy caliente (un millón de kelvin). Dicho de otra forma, la presencia de gas a 300.000 kelvin sugiere que el gas caliente se está en-

EL CIELO DE LAS NUBES

ESTE MAPA DEL GAS GALACTICO combina las observaciones por radio del hidrógeno neutro (*manchas coloreadas*) con una imagen en luz visible de la Vía Láctea (*blanco*). El mapa muestra el cielo proyectado de forma que el disco galáctico aparece en el medio; el núcleo de la galaxia ocupa el centro. Las nubes de hidrógeno de alta velocidad, como los complejos A y C, se hallan por encima y por debajo del disco.



INGRID KALLICK; *FUENTE*: BART WAKKER (*radio*), OBSERVATORIO LUND (*visible*)

friando. Junto con Blair D. Savage, de la Universidad de Wisconsin-Madison, y Kenneth Sembach, del Instituto de Ciencias del Telescopio Espacial, en Baltimore, hemos investigado esta componente caliente de las nubes de alta velocidad.

Comportamiento complejo

Tras examinar los nuevos datos, podemos ahora presentar un cuadro coherente de las nubes de alta velocidad. Comenzaremos con dos de los mayores complejos identificados de nubes de alta velocidad, el A y el C, descubiertos en 1963. El complejo A se halla a una distancia de entre 25.000 y 30.000 años-luz; está en pleno interior del halo de la galaxia. La distancia al complejo C es aún incierta: al menos unos 14.000 años-luz y no más de 45.000 sobre el plano galáctico.

Las dos nubes son deficientes en elementos pesados: una décima parte de la concentración del Sol. La cantidad de nitrógeno del complejo C es particularmente baja, apenas 1/50 de la solar. La escasez de nitrógeno indica que los elementos pesados vinieron sobre todo de las estrellas grandes, que producen menos nitrógeno, en comparación con otros elementos, que las de masa inferior. Los actuales modelos del universo joven predicen que las primeras estrellas fueron más pesadas que las actuales. Parece que el complejo C es un fósil del universo antiguo.

Brad Gibson, de la Universidad Swinburne de Melbourne, ha examinado otra región del complejo C. Encontró una concentración de elementos pesados dos veces mayor que nuestros resultados. Esta diferencia de abundancias químicas apunta a que el complejo C empieza a mezclarse con otras nubes de gas del halo galáctico que presentan una mayor concentración de elementos pesados. Además, Andrew Fox y sus colaboradores de Wisconsin se valieron de los datos del oxígeno altamente ionizado y de otros iones para probar que el gas a 300.000 kelvin del complejo C representa una interfaz entre el gas caliente y el frío. Parece que estamos observando el complejo C durante su asimilación por la galaxia.

Nubes como los complejos A y C constituyen la primera prueba directa de la caída de gas limpio en nuestra galaxia. El complejo C proporciona entre 0,1 y 0,2 masas solares de nuevo material al año; el complejo A, la mitad de eso. Estas cantidades representan entre el 10

LOS CUATRO PROCESOS QUE CONFORMAN LA GALAXIA

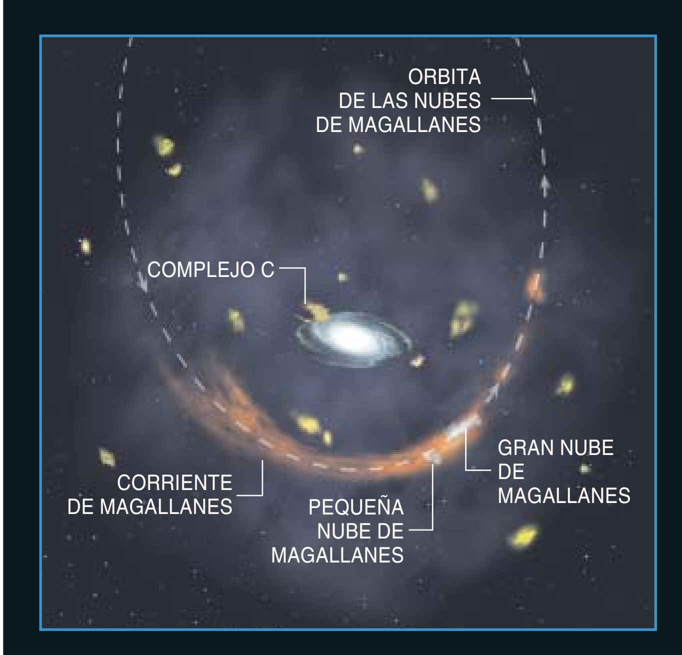
LA FUENTE GALACTICA: las nubes de velocidad intermedia representan posiblemente la rama de vuelta de un enorme ciclo de gas. Las explosiones de supernova crean burbujas de gas caliente (*azul*) que atraviesan el gas frío que las rodea (*amarillo*) y alimentan a la corona. Algunas porciones del gas caliente se enfrían y caen de nuevo al disco.



LA CAIDA DE GAS: muchas de las nubes de alta velocidad (*amarillo*) consisten en gas que se precipita hacia la Vía Láctea. El proceso de formación galáctico sigue, pues, en marcha pasados ya casi 12.000 millones de años desde que empezara. Este gas podría proporcionar nuevo combustible para la formación estelar. La observación confunde fácilmente estas nubes con las de velocidad intermedia (*naranja*).

NUBE DE VELOCIDAD INTERMEDIA
SOL
NUBE DE ALTA VELOCIDAD

CANIBALISMO GALACTICO: la Vía Láctea está arrancando gas de la Gran Nube y la Pequeña Nube de Magallanes, dos de sus galaxias satélite. A lo largo de sus órbitas discurre la corriente de Magallanes (*naranja*). Otras nubes de alta velocidad, sin relación con ésta (*amarillo*), se condensan en la corona caliente y flotan en el mismo espacio.



EL RELLENO INTERGALACTICO: la Vía Láctea y Andrómeda podrían estar sumergidas en un inmenso mar de gas intergaláctico caliente (*azul*). En él se condensan acumulaciones frías; las galaxias las capturan, dando lugar a nuevas nubes de alta velocidad que, con el tiempo, se precipitarán sobre las galaxias. Este modelo es todavía bastante incierto.



INGRID KALLICK

y el 20 por ciento del total necesario para diluir el gas de la galaxia y explicar la composición química de las estrellas. Otras nubes de alta velocidad podrían estar suministrando el resto. Pero sigue sin quedar claro del todo si la fuente original de este gas es un halo remanente (como propuso Oort), si se encuentra en el espacio intergaláctico profundo o, incluso, si perteneció a una galaxia enana digerida por la Vía Láctea.

Multitud de orígenes

Los resultados observacionales descartan tres de las hipótesis sobre el origen de los complejos A y C. La del surtidor supone que las nubes nacen en el disco. Deberían tener una composición química similar al Sol, y no es así. La de la corriente de Magallanes falla al explicar la composición del gas en cuanto a los elementos pesados. Finalmente, la de la materia oscura fracasa porque esas dos nubes de alta velocidad no se hallan en el espacio intergaláctico. No obstante, ninguna de las tres ideas está equivocada por completo. Sólo hemos de buscar el sitio donde aplicarlas.

Durante mucho tiempo, las nubes de alta velocidad postergaron a las de velocidad intermedia, menos llamativas. Varios grupos han obtenido ahora su composición química. Concuerda con la del disco galáctico. Se alzan unos 4000 años-luz sobre el plano; a esa altura es donde actuarían los surtidores. Estas dos pruebas apuntan a que son estas nubes, no las de alta velocidad, las que constituyen el flujo de vuelta de un surtidor.

Una prueba más que corrobora esta hipótesis es la detección de moléculas de hidrógeno en las nubes de velocidad intermedia. Para formarlas en el espacio se necesitan granos de polvo interestelar, lo que sólo es posible si el gas ambiente está suficientemente enriquecido. Que no se haya detectado hidrógeno molecular en el complejo C remacha esta idea. Por tanto, las nubes de velocidad intermedia serían gas reciclado del interior galáctico, mientras que las de alta velocidad contendrían sobre todo gas de su exterior.

En cuanto a la hipótesis de la corriente de Magallanes, al menos una de las nubes de alta velocidad parece realmente un fragmento desprendido de ella. Su composición química se parece bastante a la de la Pequeña Nube de Magallanes, tal y como Limin Lu y sus colaboradores de Wisconsin observaron en 1998. La nube de alta velocidad se encuentra en el brazo anterior de la corriente: lo que la arrancó de la Pequeña Nube de Magallanes, además la aceleró. Sólo les es posible a las fuerzas de marea, no a las de fricción. El descubrimiento de Lu zanja finalmente la cuestión del origen de la corriente.

Las fuerzas de fricción tienen, no obstante, su importancia. FUSE encontró oxígeno muy ionizado asociado a la corriente de Magallanes; significa que esta corriente también está inmersa en gas muy caliente. La corona galáctica se extiende mucho más lejos de lo que sostuvo Spitzer: hasta unos cientos de miles de años-luz, y no sólo hasta unos pocos miles. No es tan densa como para arrancar gas de las Nubes de Magallanes, pero una vez el gas ha sido succionado por las fuerzas de marea, la fricción con la corona sí puede desacelerarlo; así, con el tiempo, acaba por caer sobre la galaxia y contribuir a su crecimiento.

De igual modo, la hipótesis de la materia oscura, aunque incapaz de explicar los complejos A y C, tendría su lugar en un esquema más amplio. Blitz mantuvo que las nubes de alta velocidad intergalácticas pesaban entre 10 millones y 100 millones de masas solares. Todavía no se han detectado nubes de esta magnitud en grupos de galaxias cercanos que se parezcan al Grupo Local, pese a que hoy las observaciones cuentan con sensibilidad suficiente para lograrlo. Además, la hipótesis predice que la emisión en luz visible de las nubes de alta velocidad no podría captarse; sería demasiado débil. Sin embargo, casi siempre que se la ha buscado, se la ha encontrado. Por último, la teoría enseña que, si las nubes de alta velocidad son muy lejanas, deberían estar completamente ionizadas o tener una gran masa; ni lo uno ni lo otro con-

Corrientes

LA MAYOR PARTE DE LA VIA LACTEA está bien mezclada. Dos estrellas que se formaron en un mismo lugar podrían encontrarse hoy día en regiones muy diferentes del cielo. Pero en los últimos años se han encontrado grupos de estrellas que se mueven al unísono en corrientes de estrellas. Se cree que son restos arrancados de galaxias satélite por las mareas, en un proceso similar al que formó algunas de las nubes de alta velocidad. Indica, por tanto, un flujo de estrellas desde galaxias enanas hasta la Vía Láctea. No se debe confundir estas corrientes con la de Magallanes, compuesta de gas. Representan una prueba distinta del continuo crecimiento de nuestra galaxia.



Un ejemplo espectacular es el flujo de estrellas que están siendo arrancadas de la galaxia enana esferoidal Sagitario, descubierto en 1994 por Rodrigo Ibata, del Observatorio de Estrasburgo, y sus colaboradores (*véase la ilustración*). Se han encontrado después otras corrientes de estrellas gracias a los datos recogidos por el Estudio Digital Sloan de los Cielos, que cubrió sistemáticamente una gran parte del firmamento. Una de estas corrientes podría estar asociada a la galaxia enana Canis Major, según han descubierto Ibata, Nicolas Martin, de Estrasburgo también, y sus colaboradores hace unos meses. A lo largo de los últimos dos mil millones de años, esa galaxia se ha ido estirando hasta configurarse como un anillo espiral de estrellas a lo largo del plano galáctico.

RON MILLER

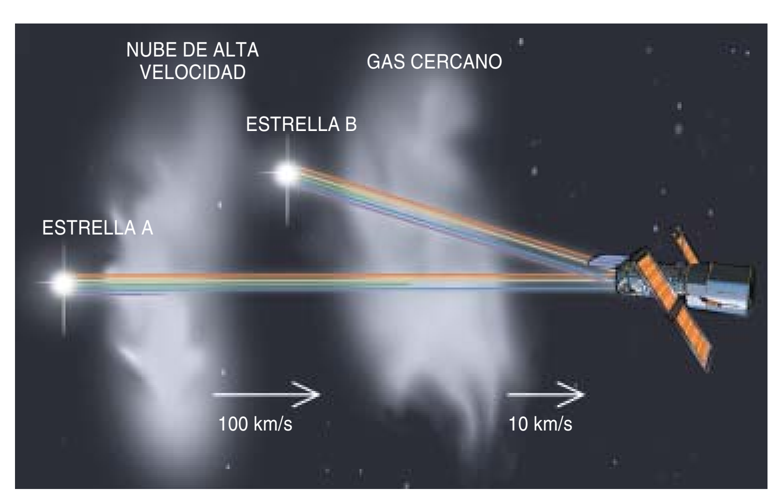
TRAS LAS NUBES

LAS NUBES DE ALTA VELOCIDAD burlaron a los astrónomos durante décadas porque se desconocía su composición y la distancia a que se encuentran. Sólo cabe medir esas propiedades por medio de las líneas de absorción. Las estrellas y las galaxias situadas detrás de las nubes de alta velocidad actúan como bombillas que brillasen a través del gas. Este absorbe algunas longitudes de onda; a partir de ahí se pueden medir las propiedades de las nubes.

Si el espectro de una estrella tiene líneas de absorción, significa que hay una nube entre la estrella y nosotros. La distancia a la estrella impone un límite superior a la distancia de la nube. Y viceversa: la carencia de líneas de absorción impone un límite inferior a la distancia de la nube. Para establecer estos límites hay que suponer que se descartan otros factores: la incertidumbre de la distancia a la estrella, elementos pesados que produzcan líneas de absorción detectables y líneas de absorción creadas por el material de la propia estrella.

Para medir la distancia a las nubes de alta velocidad, las mejores candelas son las estrellas variables RR Lyrae y las "estrellas azules de la rama horizontal". Son bastante numerosas, sus distancias se determinan con cierta precisión y sólo unas pocas de sus líneas se solapan con las líneas de las nubes. En principio, se podrían utilizar las líneas de cualquier elemento. Sin embargo, para determinar el contenido de elementos pesados es mejor emplear las líneas espectrales del oxígeno neutro y del azufre ionizado. Pertenecen a la parte ultravioleta del espectro; se requieren, por tanto, satélites apropiadamente equipados, como el Telescopio Espacial Hubble y el Explorador Espectroscópico del Ultravioleta Lejano (FUSE). En este último caso, las mejores candelas son las galaxias activas distantes, como los cuásares, porque con frecuencia tienen espectros sin rasgos y brillan más en el ultravioleta que las estrellas.

Una única estrella o galaxia puede iluminar a más de una nube de gas. Cada nube se desplaza a una velocidad diferente, por lo que absorbe la luz a longitudes de onda un poco distintas debido al efecto Doppler. Para distinguir las nubes es necesario emplear espectrómetros de alta resolución espectral, lo que a su vez requiere telescopios de gran tamaño.

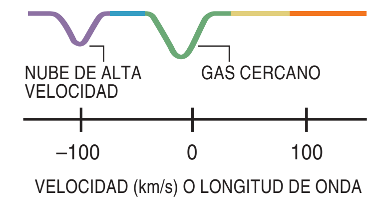


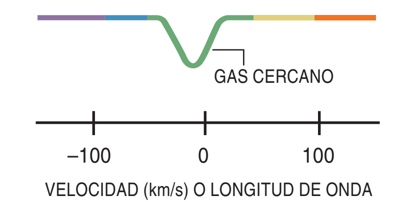
VISTA LONGITUDINAL (fuera de escala)

1. LA MAYOR DIFICULTAD en el estudio de las nubes de alta velocidad es la medida de la distancia. La mejor técnica disponible es indirecta y aproximada. Considérese una nube de alta velocidad que está entre dos estrellas, A y B. Otra nube de gas, más lenta, se encuentra entre B y nosotros.



VISTA DESDE LA TIERRA





2. CUANDO SE TOMA UN ESPECTRO de A, se observan dos líneas de absorción: se predice la existencia de dos nubes. Las líneas están en dos longitudes de onda diferentes porque la velocidad de cada una de las nubes es distinta.

3. LA ESTRELLA B tiene sólo una línea de absorción, por lo que se debe encontrar por delante de la nube de alta velocidad. Es decir, las dos estrellas, cuyas distancias se han determinado aparte por otros métodos, imponen unos límites superior e inferior a la distancia a la nube.

cuerda hoy con las observaciones. Por tanto, parece que las nubes de alta velocidad no forman parte de la población de nubes de materia oscura.

Según Robert Braun, del Observatorio Dwingeloo, y Butler Burton y Vincent de Heij, de Leiden, Andrómeda y nuestra galaxia están rodeadas de varios cientos de nubes pequeñas, compuestas principalmente de materia oscura y de gas ionizado, con una reducida fracción de hidrógeno neutro. Pesarían como mucho 10 millones de masas solares y no vagarían por el Grupo Local, sino que se encontrarían a medio millón de años-luz de las galaxias principales.

Aunque no parezca que las nubes de alta velocidad se hallen dispersas por el Grupo Local, otra clase de gas de alta velocidad sí podría estarlo. El gas muy ionizado de una

INGRID KALLICK

de las nubes de alta velocidad se encuentra lejos de la Vía Láctea. FUSE también ha descubierto oxígeno fuertemente ionizado de alta velocidad aislado, sin que lo acompañe ningún gas neutro. Todd M. Tripp, de la Universidad de Princeton, y sus colaboradores han encontrado nubes similares de gas muy caliente en otras regiones del universo. Este gas caliente quizá sea parte de un filamento que recorra el espacio intergaláctico. Filamentos así aparecen en las simulaciones de la evolución a gran escala del universo [*véase* “Vacíos espaciales”, por Evan Scannapieco, Patrick Petitjean y Tom Broadhurst; INVESTIGACIÓN Y CIENCIA, diciembre de 2002]. La cantidad total de masa que almacenan podría ser mayor que la de todas las galaxias juntas. De una reserva así podría beber la Vía Láctea para generar nuevas estrellas.

Las nubes de alta velocidad que rodean a la Vía Láctea nos recuerdan que vivimos en una galaxia que está todavía en proceso de formación y evolución. Al principio, la ceñían muchas galaxias satélites pequeñas y gas residual. A lo largo de los últimos miles de millones de años, se ha anexionado la mayoría de esos satélites y parte del gas primigenio del medio intergaláctico cercano. Sigue cayendo gas hacia la galaxia, tomando la forma de nubes de alta velocidad. Al mismo tiempo, la galaxia expulsa gas cargado de elementos pesados al halo, y puede que incluso al espacio intergaláctico.

En los próximos 10.000 millones de años, más galaxias satélites se fundirán con la Vía Láctea. Se crearán así corrientes y flujos de estrellas como los que se están descubriendo ahora en el halo. Nuestra galaxia y Andrómeda siguen un rumbo que las llevará a chocar una contra otra. No podemos predecir el aspecto de la Vía Láctea, o de lo que quede de ella, en un futuro lejano. Sólo sabemos que su formación no ha terminado aún.

Los autores

Bart P. Wakker y **Philipp Richter** observan las bandas ultravioleta y radio del espectro electromagnético. Wakker realizó su tesis doctoral en nubes de alta velocidad en la Universidad de Groninga y pasó 5 años en la Universidad de Illinois, antes de trasladarse a Wisconsin en 1995, donde trabajaría con Richter desde 1999. Richter obtuvo su doctorado en la Universidad de Bonn, donde estudió el gas difuso molecular de las Nubes de Magallanes y el halo de la Vía Láctea. Tras dejar Wisconsin en 2002, pasó al Observatorio Astrofísico de Arcetri, en Florencia. Hace poco ha regresado a Bonn.

Bibliografía complementaria

HIGH-VELOCITY CLOUDS. Bart P. Wakker y Hugo van Woerden en *Annual Review of Astronomy and Astrophysics*, vol. 35, págs. 217-266; septiembre 1997.

A CONFIRMED LOCATION IN THE GALACTIC HALO FOR THE HIGH-VELOCITY CLOUD “CHAIN A”. Hugo van Woerden, Ulrich H. Schwarz, Reynier F. Peletier, Bart P. Wakker y Peter M. W. Kalberla en *Nature*, vol. 400, págs. 138-141; 8 de julio 1999. Disponible en arXiv.org/abs/astro-ph/9907107

ACCRETION OF LOW-METALLICITY GAS BY THE MILKY WAY. Bart P. Wakker, J. Chris Howk, Blair D. Savage, Hugo van Woerden, Steve L. Tufte, Ulrich J. Schwarz, Robert Benjamin, Ronald J. Reynolds, Reynier F. Peletier y Peter M. W. Kalberla en *Nature*, vol. 402, n.º 6760, págs. 388-390; 5 de noviembre 1999.

THE FORMATION AND EVOLUTION OF THE MILKY WAY. Cristina Chiappini en *American Scientist*, vol. 89, n.º 6, págs. 506-515; noviembre-diciembre 2001.

A FAR ULTRAVIOLET SPECTROSCOPIC EXPLORER SURVEY OF MOLECULAR HYDROGEN IN INTERMEDIATE-VELOCITY CLOUDS IN THE MILKY WAY HALO. P. Richter, B. P. Wakker, B. D. Savage y K. R. Sembach en *Astrophysical Journal*, vol. 586, n.º 1, págs. 230-248; 20 de marzo 2003. arXiv.org/abs/astro-ph/0211356

HIGHLY IONIZED HIGH-VELOCITY GAS IN THE VICINITY OF THE GALAXY. K. R. Sembach, B. P. Wakker, B. D. Savage, P. Richter, M. Meade, J. M. Shull, E. B. Jenkins, G. Sonneborn y H. W. Moos en *Astrophysical Journal Supplement Series*, vol. 146, n.º 1, págs. 165-208; mayo 2003. arXiv.org/abs/astro-ph/0207562

La prueba del nueve

Los pitagóricos y los alumnos de primaria, los matemáticos árabes y los algebristas del Renacimiento han aplicado la prueba del nueve. ¿Se trata de una receta mágica?

Michel Ballieu

En la clase el ambiente es tenso. El maestro, severo pero justo, permanece en un extremo de la tarima. El niño, ante el gran pizarrón negro, gacha la cabeza, trata de esquivar la mirada de reproche que le lanza su instructor. "Vamos a ver... ¿estás seguro de que 171×231 son 39.401?", le pregunta, impaciente la voz. La respuesta es tímida. "Esto... ¿sí, señor?". El maestro estalla. "¡Pequeño cabestro! ¿No te das cuenta de que te has equivocado? ¿Te has olvidado de la prueba del nueve?" El niño, aterrorizado, guarda un silencio culpable y siente revolotear sobre sí la amenaza del castigo. La sentencia no tarda en llegar. "Abre la mano", ordena el profesor, que golpea con su larga regla de hierro la palma infantil...

Dejemos al pequeño en su doloroso trance. En nuestros días, los maestros son menos severos y la "prueba del nueve" carece de significado para los escolares, que ahora utilizan siempre la calculadora para obtener el resultado de operaciones fastidiosas. Como ya no está en boga, antes de repasar su historia diremos en qué consiste.

La prueba del nueve es un procedimiento utilizado para la detección de errores al efectuar operaciones aritméticas (adición, sustracción, multiplicación o división).

Así se explicaba la prueba del nueve para la multiplicación en un viejo manual escolar que data de 1923: "Sea que se haya de multiplicar 3587 por 286. Se traza un aspa a la derecha de la multiplicación, en la cual se anotan los respectivos restos de la división por 9 del multiplicando y del multiplicador, que son 5 y 7. En el hueco superior se anota el resto de la división por 9 del producto 35 de los dos restos 5 y 7, resto que es igual a 8. Por último se anota debajo el resto de la división por 9 del producto que se ha de comprobar. Las dos cifras escritas una sobre la otra han de ser iguales".

Tras esta receta misteriosa se ocultan razonamientos sencillos concernientes a los restos de las divisiones de números enteros naturales (0, 1, 2, 3,...). Los griegos del siglo III antes de nuestra era veían la división con resto —también llamada división euclídea— como una serie de sustracciones sucesivas. En consecuencia, el cálculo de 53 dividido por 6 consiste en las ocho restas $53 - 6 = 47$, $47 - 6 = 41$,... $11 - 6 = 5$; llegados a esta situación ya no es posible restar 6, por lo que el resto es 5. Se escribe, pues, $53 = 8 \times 6 + 5$. Este proceder pone de manifiesto que el resto de la división siempre ha de valer menos que el divisor. Cuando la división es exacta, vale 0.

¿Cómo se comportan los restos de los diversos términos de una operación aritmética al dividirlos por un mismo divisor, por ejemplo el 7? Examinemos la suma de 34 y 38. Empezamos dividiendo estos números por 7. Se tiene $34 = 7 \times 4 + 6$ y $38 = 7 \times 5 + 3$. Los restos respectivos son 6 y 3, ambos menores que 7. Al sumar miembro a miembro las igualdades anteriores y sacar factor común 7 en la suma de los segundos miembros tendremos $34 + 38 = 7 \times (4 + 5) + (3 + 6)$. La suma de los restos es $3 + 6 = 9$, mayor que 7; hay, pues, que restar 7 una vez más para obtener el resto de dividir por 7 el resultado de la suma. Con toda generalidad, el resto de la suma de dos enteros es igual al resto de la suma de sus respectivos restos: en efecto, 2 es tanto el resto de la división por 7 tanto de $6 + 3$ como de $34 + 38 = 72$. De ahí que una posible verificación de una suma consista en comprobar que, para un divisor dado, el resto de la suma es igual al resto de la suma de los restos de cada uno de los sumandos.

Basta sumar las cifras

La multiplicación consiste en una serie de adiciones sucesivas, por lo que será válida para ella la misma regla. El producto de aquellos dos números se expresa, una vez separado el factor 7, como $34 \times 38 = (7 \times 4 + 6) \times (7 \times 5 + 3) = 7 \times (4 \times 7 \times 5 + 4 \times 3 + 6 \times 5) + (6 \times 3)$. En este caso, como antes, se ha de tomar el resto de la división por 7 del producto de los dos restos; en general: el resto del producto de dos en-

teros es igual al resto del producto de sus restos. El resto de $6 \times 3 = 18$ al ser dividido por 7 vale 4, que es también el resto de $34 \times 38 = 1292$ al dividirlo por 7.

Carl Friedrich Gauss, que fue llamado *Princeps mathematicorum* por sus coetáneos, introdujo en su célebre obra *Disquisitiones artihmeticae*, publicada en 1801, el formalismo de los números congruentes, de las clases de restos y de los módulos, merced al cual es posible reformular con elegancia los asertos precedentes. Se dice que a y b son congruentes módulo m si tienen el mismo resto al dividirlos por m; se escribe entonces $a \equiv b \pmod{m}$ y se lee "a es congruente a b módulo m". Este resto, o residuo, se denota a mod m. Tenemos así que 9 mod 7 = 23 mod 7 = 2. Con estas notaciones, las dos propiedades en que se funda la prueba del nueve se expresan de forma muy concisa:

$(a + b) \bmod m = (a \bmod m + b \bmod m) \bmod m$, y

$(a \times b) \bmod m = (a \bmod m \times b \bmod m) \bmod m$.

La prueba del nueve clásica se obtiene tomando los restos de las divisiones por 9, lo que equivale a calcular módulo 9.

Dos observaciones. La primera: es obligado ser prudente. La prueba del nueve no da garantía plena de la validez del resultado. En efecto, dos números congruentes módulo 9 darán el mismo resto al dividirlos por 9. Si al efectuar la operación se comete un error que sea múltiplo de 9, como sucede en el caso de que se permuten dos cifras, la prueba del nueve no lo señalará, porque el resultado correcto y el erróneo serán congruentes módulo 9. Por ejemplo, al dividir por 9 los números 6271, 6172 o 6721 se obtiene, en todos los casos, un resto 7. La prueba del nueve tiene más de procedimiento para la detección de errores que de método de demostración. Hay docentes que en nuestros días se valen de la prueba del nueve para ejemplificar las nociones de condición necesaria y de condición suficiente. Si una operación aritmética ha sido bien efectuada, necesariamente habrá de cumplir la prueba del nueve. El fallo en la prueba es suficiente para tener la certeza de que se ha cometido un error. Pero no es suficiente que la operación supere la prueba para tener la certeza de que el cálculo ha sido correctamente efectuado.

Por otra parte, aunque en la comprobación por congruencia se puede utilizar cualquier número, la prueba del nueve ha sido la más empleada debido a la gran facilidad con que se obtiene el resto de una división por 9: basta sumar todas las cifras del número que se ha de dividir, y repetir este proceso hasta que se obtenga un número menor que 9, que será el resto. En efecto, al dividir por 9 cualquier potencia de 10 el resto es igual a 1. La propiedad del producto de congruencias nos hace ver que el resto del término correspondiente a cada una de las potencias de 10 de que consta un número es igual al coeficiente de dicha potencia. Al sumar módulo 9 estos coeficientes se obtiene el resto de dividir el número por 9. Todos los escolares conocían esta regla "allá en los buenos tiempos".

Mónadas y pitmenos

La historia de la prueba del nueve se remonta a los pitagóricos, del siglo V antes de nuestra era.

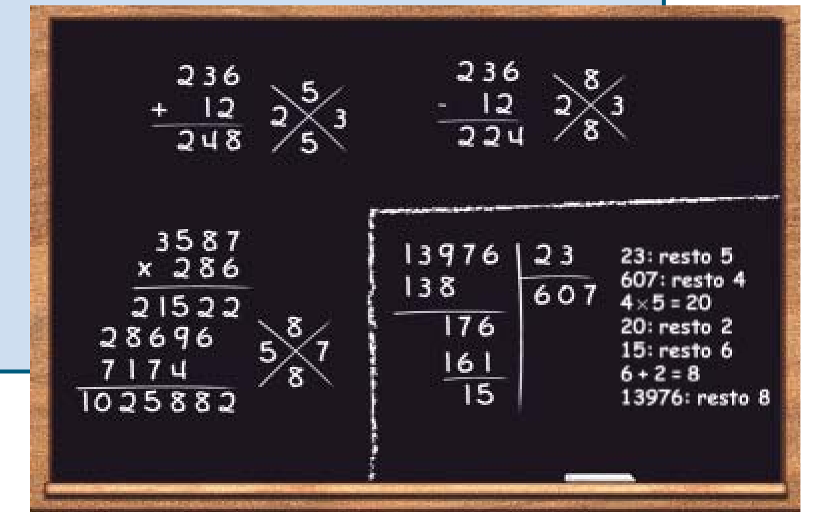
Yámblico, filósofo neoplatónico nacido en Calcis, en la actual Siria, escribió a principios del siglo IV de nuestra era nueve libros consagrados a la secta pitagórica. Uno de ellos, titulado *Sobre la introducción de la aritmética de Nicómaco de Gerasa*, recopila comentarios relativos a una obra de este neoplatónico del siglo I d.C. Los pitagóricos atribuían virtudes especiales a las potencias de 10, a las que denomina-

Así se utiliza la prueba del nueve

La prueba del nueve es un procedimiento de comprobación del resultado de una operación y se funda en la comparación de los restos obtenidos al dividir por 9 los operandos. Su facilidad de empleo se debe a que, para obtener dichos restos, basta sumar las cifras que componen los números. Ilustremos el procedimiento con la suma de la pizarra de la derecha. El resto de la división por 9 de 236 es 2, que hallamos así: 2 + 3 + 6 = 11, y 1 + + 1 = 2. De igual forma, el resto de 12 es 3. Estos restos se trasladan a los lados derecho e izquierdo del aspa, y se suman: 3 + 2 = 5. Se toma nuevamente el resto de la división por 9 de este número, que es, claro está, 5, y se escribe en lo alto del aspa. A continuación se calcula el resto del resultado de la operación, que resulta también igual a 5 (2 + 4 + 8 = 14, y 1 + 4 = 5), y lo escribimos en el hueco restante. Se comparan los números arriba y abajo del aspa: como los dos restos son iguales, es muy posible que el resultado sea correcto.

La sustracción funciona de igual manera, con la salvedad de que ahora se han de sustraer los restos de los operandos (y tener en cuenta que 2 – 3 = –1 = 8 mod 9); y para la multiplicación, se los multiplica.

En la prueba de la división hay que tener en cuenta el resto de la operación. Sea dividir 13.976 entre 23. Se calcula el resto de 23, que es 5. El resto de 607 vale 4. Se multiplican uno por otro estos dos restos y se toma el resto del producto, que es 2, al cual se le suma el resto de 15, que vale 6. El total es 8. Para terminar se compara este resto con el resto de 13.976. También vale 8: el cálculo parece correcto.



La prueba del 9 en al-Khwarizmi y en Leonardo de Pisa

La obra de aritmética de al-Khwarizmi se redactó a principios del siglo IX, pero el único manuscrito conocido hasta la fecha es una traducción latina que data del siglo XIII. El fragmento siguiente explica el método de la prueba del nueve:

"...Y cuando quieras multiplicar un número cualquiera por otro número y quieras hacer la prueba (...), divide por 9 el número que has multiplicado, y lo que sobre de este lado de 9, guárdalo. De igual manera, divide el segundo número por 9, y lo que sobre de este lado de 9, guárdalo. A continuación, multiplica lo que te sobró la primera vez por lo que te sobró la segunda, y elimina de este resultado todo lo que, reunido, dé 9, si es que lo hay, y si no lo hay, lo que quede será la marca. Pero si hay 9, echa fuera 9 y guarda lo que quede, y entiende que ésa será la marca. Después, multiplica uno por otro los números que has de multiplicar y divide el producto por 9, y si lo que queda es lo mismo que lo que te he dicho al respecto de la marca, sabe que has acertado. Si no es lo mismo, comprende que te has equivocado".

El capítulo V del *Liber abaci* de Leonardo Fibonacci contiene un ejemplo y una demostración de la prueba del nueve. Fibonacci menciona asimismo la prueba por 7 y la prueba por 11. El ejemplo siguiente estudia la división de 13.976 entre 23, cuyo resultado exacto vale 607 + 15/23:

"Prueba de la división escrita arriba. Si se quiere comprobar la división (...) por el resto por 9, se toma el resto de 13.976, que vale 8, y se guarda a un lado. Y, de nuevo, se toma el resto del número resultante, a saber, de 607, que vale 4, y se multiplica por el resto de 23, que vale 5, lo que da 20, del que se toma el resto, que vale 2, y se suma con el 15 que está sobre la barra de fracción de 23, lo que da 17, cuyo resto vale 8, lo mismo que habíamos guardado a un lado más arriba. Por ejemplo, puesto que del producto del divisor por el número resultante se obtiene el número dividido, si nosotros multiplicamos el resto del divisor por el resto del número resultante, nos dará el resto del número dividido. Pero del número dividido por 23 se ha restado 15, que extraído de 13.976 dejan 13.961, que dividido por 23 hace 607 (...). Por eso, si multiplicamos el resto de 607, que vale 4, por el resto de 23, que vale 5, resulta 20, cuyo resto, a saber, 2, es el resto de 13.961, que sumado al resto del 15 que está encima, y que vale 6, da un total de 8, es decir, el resto de 13.976, y eso es lo que yo he querido demostrar. Desde luego, las multiplicaciones, sustracciones o divisiones de los números pueden ser comprobados de distinto modo mediante ciertos restos, como el de 7 y los de todos los números sordos [*números primos*] existentes, como 11, o 13, y así sucesivamente. Esta teoría, que en nuestra opinión es conveniente, la demostraremos en lo que sigue".

La prueba por nueve de la división explicada por Leonardo Fibonacci se expresa en términos más concisos: 13.976 mod 9 = (23 mod 9 $\times$ 607 mod 9) mod 9 = 8.

ban mónadas (término que procede de una raíz griega que significa "unidad"). Los griegos no poseían el sistema de numeración posicional, y llamaban mónadas de primer orden a las unidades de unidades, mónadas de segundo orden a las unidades de decenas, y así sucesivamente.

Yámblico enuncia una propiedad sorprendente de las sumas de mónadas: tomemos tres números consecutivos, de los que el mayor sea divisible por 3, y sumémoslos. Calculemos después la suma de las mónadas de todos los órdenes que haya en el resultado, y repitamos el proceso: siempre encontraremos 6. Verifiquemos que así ocurre con los números 871, 872 y 873 (que es divisible por 3). Suman 2616. Esta suma posee seis mónadas del primer orden, una del segundo orden, etc. Su suma, 15, posee cinco mónadas de orden uno y una de orden 2, ¡y su suma es 6!

No hay misterio alguno. Sumar las mónadas equivale, en nuestra numeración posicional, a sumar las cifras del número, lo mismo que en la prueba del nueve. La propiedad enunciada por Yámblico es evidente si escribimos los tres enteros consecutivos en la forma $(3a + 1)$, $(3a + 2)$ y $(3a + 3)$, cuya suma vale $9a + 6$. Evidentemente, al dividir por 9 esta expresión, el resto es 6. Este resto, como hemos visto, se obtiene al efectuar la suma de sus cifras.

Predicciones arriesgadas

Hallamos semejanzas con el procedimiento de la prueba del nueve en las obras de San Hipólito de Roma, que fue obispo en el paso del siglo II al III y murió mártir. Una de sus obras, la *Refutación de todas las herejías*, contiene un capítulo dedicado a las "Predicciones por medio del cálculo." El término *pitmeno* empleado en el párrafo siguiente, y que encontramos también en otros autores, como Platón o Apolonio, representa el número de mónadas; corresponde a la noción de cifra en nuestra numeración:

“Quienes pretenden hacer profecías por medio de cálculos y de números (...) afirman que para todo número hay un pitmeno. En el caso de los millares, hay tantas mónadas como millares: por ejemplo, el pitmeno de seis mil, seis mónadas (...). Y en el caso de las centenas, tantas mónadas serán sus pitmenos como centenas haya (...). Y en el caso de las mónadas, las propias mónadas son los pitmenos, por ejemplo, de nueve, nueve; de ocho, ocho (...). Debemos actuar de igual manera en el caso de las (...) palabras, porque cada letra corresponde a un cierto número; por ejemplo, la letra ν, que representa el número cincuenta, tiene al 5 por pitmeno (...)”.

Hay que saber que el sistema de numeración griego era alfabético. Las letras α, β, γ, δ, ε, ... , ι, κ, λ, μ, ν, ..., ρ, σ, τ, υ, φ, ... representaban, respectivamente, a los números 1, 2, 3, 4, 5, ..., 10, 20, 30, 40, 50, ..., 100, 200, 300, 400, 500, ... La letra ω representa el 800. El autor latino prosigue entonces:

“Las letras del nombre de Héctor (Εκτωρ) son cinco en número. Sus pitmenos son 5, 2, 3, 8, 1; si los sumamos resulta 19. De nuevo, el pitmeno de 10 vale 1 y el de 9, 9. Si los sumamos se obtiene 10, cuyo pitmeno vale 1. El nombre de Héctor (...) ha dado una raíz [una unidad]. Sin embargo, sería más fácil (...) dividir la suma de los pitmenos entre nueve, y lo que reste, será la raíz. En el caso de Héctor, si divido por 9, el resto es 1”.

San Hipólito, que condena esas “predicciones”, da múltiples ejemplos, y explica cómo no faltan quienes saquen conclusiones. Según el mismo cálculo, a Aquiles le son atribuidas cuatro mónadas. ¡Por lo tanto, vence a Héctor!

Las propiedades de las mónadas y de los pitmenos no pasan de ser curiosidades, pero la prueba del nueve en cuanto procedimiento de detección de errores nos ha llegado de los matemáticos árabes, quienes, tal vez, la recibieron de la India, como atestigua ibn Sina, más conocido por el nombre de Avicena (980-1037).

El matemático árabe Muhammad ibn Musa al-Khwarizmi (alrededor de 780-850) alude a la prueba del nueve en su tratado de aritmética, compuesto en Bagdad a principios del siglo IX. al-Khwarizmi significa “que procede de Khwarizm”, una región al sur del mar de Aral. La deformación de este gentilicio en los tiempos del Renacimiento dio lugar al nombre “algoritmo”, con el que hoy es designado un sistema finito de operaciones que permite la resolución de un problema aritmético dado. La prueba del nueve constituye un perfecto ejemplo de algoritmo de detección de errores.

Otra obra de al-Khwarizmi, que lleva por título *al-Kitab almuhtasar fi hisab al-yabr u-al-muqabala* es el origen de la palabra “álgebra.” *Yabr* y *muqabala* designan las operaciones mediante las cuales las ecuaciones de grado uno o dos son reducidas a alguna de las formas canónicas que al-Khwarizmi enseña a resolver acto seguido. El tratado de aritmética de al-Khwarizmi es la primera obra conocida donde se exponen el sistema de notación posicional decimal y los procedimientos para efectuar las operaciones en él; incluye ejemplos de la prueba del nueve. Las pruebas del nueve, del siete o del once aparecen también en las obras de otros tratadistas árabes: el matemático al-Kashi, fallecido en 1429, utiliza la expresión *al-mizan bi tis’a*, que significa “balanza del nueve”. Ibn al-Banna, muerto en 1321, y al-Qalasadi, muerto en 1486, empleaban la expresión *al-ijtibar bi tis’a*, “sometido a la prueba del nueve”.

La travesía del Mediterráneo

El inevitable Leonardo Fibonacci, llamado Leonardo de Pisa, introdujo este método en Occidente. Lo poco que se sabe de él procede de los datos que nos proporciona la más célebre de sus obras, el *Liber abaci*, compuesta en 1202 y revisada en 1228. Cuenta que su padre, que era contable de los mercaderes de Pisa, le llevó a la aduana de Bejaia, en Kabilia. El joven Leonardo descubrió allí el cálculo con las cifras indias, cuyo empleo no deja a partir de entonces de preconizar en Occidente. Los siete primeros capítulos del *Liber abaci* están dedicados al cálculo mediante el sistema de numeración posicional indoárabe; en el capítulo V, Fibonacci expone una demostración de la prueba del nueve e indica que también es posible utilizar la prueba por 7 y por 11.



EL PRIMER FOLIO de un manuscrito del *Liber abaci*, de Leonardo Fibonacci, compuesto en 1228, en el que el autor refiere su encuentro con la numeración hindú y prodiga los consejos para la utilización de su obra. La prueba del nueve se expone en el capítulo V.

La prueba del nueve llegó así a Occidente, donde se perpetuó. La encontramos sobre todo en la *Summa* de Luca Pacioli (1445-1514). Pacioli, que se refiere a Fibonacci, utiliza las pruebas por nueve y por siete, pero no da reglas prácticas para hallar el resto en el caso de esta última. Tras Pacioli, la prueba del nueve queda sólidamente implantada en Occidente. Robert Recorde la emplea en su obra *The Ground of Arts*, publicada en Londres hacia 1542. En 1556, Pierre Forcadel enuncia un criterio de divisibilidad por 7 y lo utiliza para la prueba del siete en su *L'arithmétique de P. Forcadel de Béziers*.

Posteriormente, los matemáticos, para quienes la prueba del nueve es tema agotado, dirigen sus investigaciones hacia los criterios de divisibilidad por enteros cualesquiera. Pascal, Gorgonne, Fontenelle, d'Alembert o Lagrange, entre otros, se aplicaron a la cuestión.

La prueba del nueve sobrevivió en las escuelas: en 1872, por ejemplo, el *Traité d'arithmétique* de Joseph Bertrand, que contenía las materias exigidas para la admisión en la Escuela Politécnica, exponía las pruebas del once y del nueve. Esta última figura en lugar destacado en los manuales escolares de comienzos del siglo XX. En nuestros días quedan todavía algunos maestros que la enseñan, a pesar de que no figura en ningún programa. Sin duda por nostalgia de la escuela primaria...

El autor

Michel Ballieu es profesor del Centro de investigaciones sobre la enseñanza de las matemáticas de Nivelles, en Bélgica.

Bibliografía complementaria

MUHAMMAD IBN MUSA AL-KHWARIZMI. LE CALCUL INDIEN. A. Allard. Blanchard, París, 1992.

LE *LIBER ABACI*, CE QU'ON Y TROUVE EFFECTIVEMENT. M. Ballieu en *Nouvelles tendances en histoire et philosophie des sciences*. Académie Royale des Sciences, des Lettres et des Beaux-Arts de Belgique; Bruselas, 1993.

En torno a una econo

Es muy grande el interés que suscitan las pilas de combustible.
Sin embargo, no resulta tarea fácil desarrollar una economía que dependa del hidrógeno



mía del hidrógeno

Matthew L. Wald



SLIM FILMS

En otoño de 2003, a los pocos meses de que el presidente Bush anunciara un programa de investigación de casi 1500 millones de euros para el desarrollo de un vehículo que contaminara menos la atmósfera y redujera la dependencia del petróleo importado, la empresa Toyota presentó en Washington dos propuestas en ese sentido. En una de ellas, un sedán híbrido ya comercializado, el motor de explosión de gasolina estándar se combinaba con un motor eléctrico alimentado por baterías. Podía recorrer unos 20 kilómetros por litro y emitía la mitad de dióxido de carbono que un coche de tipo medio. El otro automóvil, un todoterreno, utilizaba hidrógeno en pilas de combustible para propulsar el motor eléctrico y sólo desprendía algo de calor y vapor de agua. ¿Cuál de ellos respetaba más el entorno?

Hallar la respuesta correcta repercutiría notablemente en la inversión dedicada a investigación, en las subvenciones que el gobierno concede para estimular el desarrollo de técnicas que independicen de la gasolina y, por fin, en el medio. Sin embargo, la resolución del dilema encierra mayor dificultad de lo que parece a primera vista. Robert Wimmer, responsable de investigación en cuestiones técnicas y legales de Toyota, y otros creen que ambos vehículos resultan más o menos equivalentes.

Los expertos van dejando de evaluar el impacto ambiental a partir de la emisión de contaminantes, para abordarlo desde una visión más global de su incidencia en el entorno: consideran el efecto combinado del combustible y el funcionamiento del vehículo. Desde esta perspectiva más amplia, incluso las ensalzadas ventajas de las pilas de hidrógeno, eficacia y limpieza, no resultan tan determinantes como podría creerse. Visto así, alcanzar el nivel de los coches híbridos supone ya un logro apreciable. El Departamento de Energía estadounidense (DOE) declara que, en ciertos casos, la propulsión por pilas de combustible puede aumentar de forma notable las emisiones de dióxido de carbono y otros contaminantes. En cuanto a la prontitud de la aplicación, el vehículo híbrido se muestra superior: se encuentra ya en el mercado y, por tanto, se puede emplear para reducir la

contaminación desde ahora mismo. Los coches con pilas de combustible, en cambio, aún tardarán en llegar. Se esperan en fechas comparables a las del vuelo tripulado a Marte de la NASA y con un nivel de verosimilitud parecido.

Este sorprendente retraso revela las incógnitas y obstáculos que jalonan la carrera hacia una economía basada en el hidrógeno, en la que la mayor parte de la energía se invertiría en la producción de hidrógeno, que luego se dirigiría a una pila de combustible para generar electricidad. Se han puesto muchas esperanzas en el desarrollo de pilas de combustible y en la posibilidad de una economía del hidrógeno más limpia, que afectaría no sólo al transporte, sino también a la distribución de energía a viviendas y otros edificios. El pasado noviembre, el secretario estadounidense de Energía, Spencer Abraham, manifestó ante homólogos de 14 países y de la Unión Europea que el hidrógeno revolucionaría el mundo en que vivimos. Sólo en los Estados Unidos, los vehículos motorizados (más de 200 millones) consumen cerca de dos tercios de los 20 millones de barriles de petróleo que se gastan a diario allí.

Pero el hidrógeno tiene un coste económico y ambiental. La pila de hidrógeno combustible cuesta, por unidad de potencia producida, casi cien veces más que un motor de explosión. Para ser competitivos en precio, el watt habría que venderse a unos cuatro céntimos de euro; sin embargo, se está ofreciendo a casi tres euros y medio, recuerda Tim R. Dawsey, de la compañía Eastman Chemical, fabricante de polímeros para pilas de combustible. En cuanto a la unidad de energía utilizable, el hidrógeno también resulta unas cinco veces más caro que la gasolina. El coste ambiental se centra en la contaminación derivada de la energía que se requiere para producir hidrógeno. Y aun cuando esa energía provenga de una fuente renovable, solar o eólica, puede que tenga aplicaciones más inofensivas para el entorno que la producción de hidrógeno. La distribución y el almacenamiento del hidrógeno, así como la seguridad en su manipulación —el gas más ligero del universo— plantean otras dificultades técnicas y de infraestructura. Cualquier propuesta viable en favor de la economía del hidrógeno deberá abordar todas estas cuestiones.

Producción de hidrógeno

Las pilas de combustible de hidrógeno ofrecen dos atractivos evidentes. En primer lugar, no contaminan [*véase* "El vehículo del cambio", por Lawrence D. Burns, J. Byron McCormick y Christopher E. Borroni-Bird; INVESTIGACIÓN Y CIENCIA, diciembre de 2002]. En segundo lugar, el hidrógeno se obtiene a partir de multitud de fuentes. De hecho, no puede considerarse un combustible, en el sentido común del término. Por combustible entendemos una sustancia que se encuentra en la naturaleza, como el carbón, o que se refina a partir de un producto natural, como el gasóleo (que deriva del petróleo), y se quema luego para obtener energía. En nuestro planeta, el hidrógeno puro (H_2) no se encuentra en la naturaleza; para obtenerlo, debemos recurrir a tratamientos un tanto alambicados. Se trata, pues, de un vector energético, un agente que almacena y transporta energía desde una fuente original hasta una máquina generadora de electricidad. En palabras de David K. Garman, secretario adjunto de la oficina de Eficiencia Energética y Energías Renovables del DOE, "la belleza del hidrógeno reside en la diversidad de combustibles que admite". No obstante, cada fuente tiene sus inconvenientes.

El hidrógeno se obtiene, por ejemplo, mediante el proceso de electrólisis, que se sirve de corriente eléctrica para descomponer la molécula de agua. La electricidad suministrada provendría de células solares, generadores eólicos, centrales hidroeléctricas o reactores nucleares de próxima generación, más seguros [*véase* "Nueva generación de la energía nuclear", por James A. Lake, Ralph G. Bennett y John F. Kotek, INVESTIGACIÓN Y CIENCIA, marzo de 2002]. También se está investigando el uso de microorganismos para transformar en hidrógeno la biomasa, que incluiría partes de las cosechas que hoy carecen de valor económico. En febrero de este año de 2004, la Universidad de Minnesota y la de Patras presentaron un reactor químico que genera hidrógeno a partir de una mezcla de etanol y agua. Pese a su atractivo, todas estas técnicas por ahora resultan intransferibles al circuito comercial; lo más probable es que sigan así muchos años.

Se puede producir hidrógeno mediante la electricidad generada por la combustión de carbón, la fuente de energía más barata en muchas regiones. Pero si el carbón fuera el ingrediente básico de la economía del hidrógeno, las emisiones de dióxido de carbono promoverían el calentamiento global hasta niveles muy peligrosos.

Otra posible fuente de energía para la producción de hidrógeno es el metano contenido en el gas natural (así como el metanol u otro combustible hidrocarburo). El gas natural, al reaccionar con vapor de agua, produce hidrógeno y dióxido de carbono. Sin embargo, empleado para la carga de pilas de combustible, el gas natural perdería su utilización industrial más ventajosa hoy día: la combustión en turbinas

Resumen/Economía del hidrógeno

- Las pilas de hidrógeno ofrecen un rendimiento que dobla el de los motores de explosión usuales. A diferencia de éstos, las pilas de combustible sólo desprenden vapor de agua y calor.
- El hidrógeno puro no se encuentra en la naturaleza. Su producción depende de las fuentes de energía actuales: costosas y de accesibilidad restringida (electrólisis mediante energías renovables como la solar, eólica o hidroeléctrica) o muy contaminantes (debido a las emisiones de gases invernadero que derivan del uso de carbón u otros combustibles fósiles).
- Si bien el hidrógeno no constituye la panacea, sí puede resultar apropiado para ciertas aplicaciones. Entre ellas tal vez no figure el transporte.





FRENTE A FRENTE: Si se considera el impacto ambiental del ciclo de vida completo de un determinado combustible, el Toyota Prius (*derecha*), automóvil híbrido que combina un motor de gasolina con un motor eléctrico, resulta más respetuoso con el entorno que el "todoterreno" experimental de la misma compañía, que opera con pila de hidrógeno (*izquierda*).



de ciclo combinado y alto rendimiento para generar electricidad. A su vez, esto podría conducir a un mayor consumo de carbón, con todos los inconvenientes que ello conlleva. Las centrales de ciclo combinado convierten en electricidad el 60 por ciento del calor de la combustión del gas natural, mientras que la planta de carbón sólo aprovecha el 33 por ciento. Además, el gas natural, al quemarse, produce muy poco más de la mitad de dióxido de carbono por unidad de calor que el carbón: 210 kilogramos por millón de kilocalorías frente a 380. (La gasolina se sitúa entre el carbón y el gas natural, con 282 kilogramos de dióxido de carbono por millón de kilocalorías.) En resumen, en términos ambientales, utilizar el gas natural para producir la electricidad que se suministra a la red y así ahorrar carbón resulta mejor que transformarlo en hidrógeno para gastar menos gasolina.

Otros dos combustibles proporcionarían hidrógeno por reformado ("reforming") con vapor: el petróleo de Venezuela o del golfo Pérsico y, una vez más, el carbón extraído de las minas de los montes Apalaches. En este caso, para evitar que el dióxido de carbono liberado en la obtención de hidrógeno a partir de combustibles fósiles contribuyera al temido cambio climático, el carbono debería secuestrarse, para que no llegara a la atmósfera. A buen seguro, ese proceso entrañaría menos complejidad que secuestrar el gas carbónico que emiten millones de tubos de escape.

Según afirma Michael Wang, del Laboratorio Nacional de Argonne, lo mejor desde una perspectiva global sería acabar con las centrales térmicas de carbón, debido al elevado contenido en carbono de este combustible. En EE.UU., algo más de la mitad de los kilowatts-hora producidos provienen del carbón y alrededor del 20 por ciento del gas natural. El resto procede mayoritariamente de fuentes de energía exentas de carbono, sobre todo nucleares e hidroeléctricas. Por tanto, en la sustitución de las centrales consumidoras de carbón probablemente se tardaría decenios.

En cualquier caso, si el suministro de hidrógeno aumentara de repente, las pilas de combustible ni siquiera constituirían la mejor utilización del gas. En un artículo reciente Reuel Shinnar, profesor de ingeniería química de la Universidad Municipal de Nueva York, examinaba las alternativas de producción de combustibles y energía. En lugar de utilizar el hidrógeno como combustible, sugería algo mucho más sencillo: emplearlo para mejorar la calidad de los combustibles fósiles a través de las técnicas de hidrocraqueo (craqueo en presencia de hidrógeno) e hidrotratamiento (purificación de hidrocarburos mediante el uso de hidrógeno). Según sus cálculos, de ese modo se ahorrarían, sólo en Estados Unidos, tres millones de barriles de crudo diarios. Ambos procesos se inician con moléculas de crudo que no se pueden aprovechar para la gasolina, por tener un tamaño excesivo y una razón de carbono a hidrógeno demasiado rica en carbono. Si bien costosos, estos tratamientos todavía resultan rentables para las refinerías, pues permiten separar componentes sólo aptos para el asfalto, combustibles de calderas y otros productos de bajo valor, y convertirlos en gasolina.

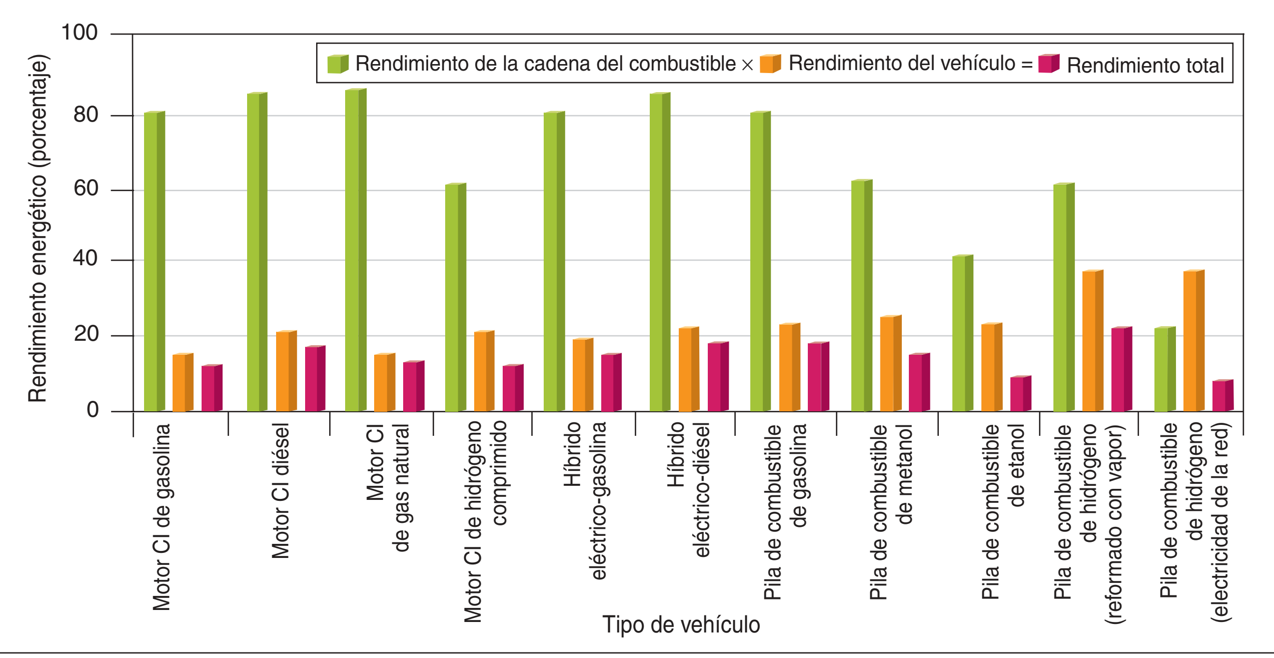
Costes de conversión

Aunque la producción de hidrógeno no sea limpia ni barata, ¿podría el impresionante rendimiento de las pilas de combustible compensar estos aspectos negativos? Una vez más, la respuesta no es sencilla.

Un kilogramo de hidrógeno contiene aproximadamente la misma energía que cuatro litros de gasolina normal sin plomo: es decir, uno y otro liberarían al arder la misma cantidad de calor. Sin embargo, la pila de hidrógeno y el motor de explosión difieren en la capacidad de extraer trabajo útil de esa energía del combustible. En el motor, la mayor parte de la energía se evacua por el tubo de escape en forma de calor; además, en el interior también se pierde energía por rozamientos. Defensores y detractores coinciden en que, en números redondos, una pila de combustible consigue de un kilogramo de hidrógeno un trabajo doble del que un motor desarrolla a partir de cuatro litros de gasolina. (En aplicaciones fijas, como una planta instalada en

RENDIMIENTO ENERGETICO "DEL POZO A LAS RUEDAS"

EL RENDIMIENTO energético total comprende no sólo el funcionamiento del vehículo, sino también la energía requerida para producir el combustible. El proceso que abarca la extracción del crudo, su refinación y el transporte de la gasolina hasta los surtidores que abastecen los motores de explosión o combustión interna (CI) ofrece un rendimiento mayor que el de la producción de hidrógeno para las pilas de combustible.



LUCY READING; *FUENTE:* DEPARTAMENTO DE ENERGIA

el sótano de una casa que extraiga hidrógeno del gas natural y lo transforme en electricidad para abastecer la vivienda, el rendimiento sería incluso mayor si, por ejemplo, se aprovechara el calor desprendido por la pila para calentar el agua.)

Existe un procedimiento que permite evaluar cuál sería la mejor forma de aprovechar un combustible. Se trata del método "del pozo a las ruedas", que compara el rendimiento energético de todas las técnicas que existen para hacer girar las ruedas de un vehículo. El elemento básico de este análisis es el "rendimiento de la conversión", que mide la pérdida de potencial energético del combustible que se produce en cada una de las etapas de la cadena de energía, desde el bombeo del crudo ("del pozo...") hasta su combustión en el motor ("...a las ruedas").

La primera parte de esta evaluación energética corresponde a la etapa "del pozo al depósito". Se centra en los costes de fabricación y entrega del combustible. Cuando el gas natural se craquea para obtener hidrógeno, se pierde alrededor del 40 por ciento de la energía. Tomar la electricidad de la red de distribución para producir hidrógeno por electrólisis del agua acarrea una pérdida del 78 por ciento. (Pese a su bajo rendimiento, la electrólisis probablemente predominará en los primeros estadios de una economía del hidrógeno debido a su comodidad: el hidrógeno se produce allá donde se necesita y, por tanto, se evitan los problemas de transporte.) En cambio, bombear crudo desde el pozo, transportarlo a la refinería, convertirlo en gasolina y llevarlo a una estación de servicio origina pérdidas en torno al 21 por ciento del potencial energético. En la obtención de gas natural y su compresión en un depósito se pierde solamente un 15 por ciento.

La segunda parte, "del depósito a las ruedas", consiste en evaluar la fracción de la energía contenida en el depósito del vehículo que realmente se utiliza para mover las ruedas. En un motor de gasolina estándar se pierde el 85 por ciento de esa energía. En suma, el sistema entero —"del pozo al depósito" y "del depósito a las ruedas"— sufre una pérdida total del 88 por ciento.

La pila de combustible trasfiere alrededor del 37 por ciento de la energía del hidrógeno a las ruedas. Si el hidrógeno se obtiene del gas natural reformado con vapor, la pérdida "del pozo a las ruedas" es del 78 por ciento. Si se obtiene mediante electrólisis con electricidad generada por carbón, la pérdida "del pozo (en este caso, una mina) al depósito" asciende al 78 por ciento; sumado al 63 por ciento que se pierde en la pila, la pérdida global se convierte en un 92 por ciento.

Wally Rippel, ingeniero de la compañía AeroVironment, que participó en el desarrollo del coche eléctrico EV-1 de General Motors y del avión eléctrico solar Helios de la NASA, ofrece un enfoque alternativo. Según sus cálculos, en un coche propulsado por un motor eléctrico, el kilowatt-hora utilizado para recargar las baterías le permitirá avanzar el triple de la distancia que recorrería si el mismo kilowatt-hora se utilizara para producir hidrógeno para una pila de combustible.

Estos hechos apoyan la tesis de los que recomiendan no utilizar electricidad para producir hidrógeno, para luego regenerar esa misma elec-

tricidad en una pila de combustible bajo el capó del coche. Sin embargo, existe una razón de peso para recurrir a las conversiones múltiples de bajo rendimiento: son más baratas. Desde el punto de vista económico no resultan tan "ineficientes". Y si tenemos en cuenta que el dinero ha sido hasta ahora el agente que ha configurado el mercado energético, aunque las pilas de hidrógeno desperdicien recursos, si las diferencias en el coste de la energía son tan enormes, ¿por qué no mudarnos a un sistema que nos permita acudir allí donde la energía sea más barata?

Walter Schroeder, de Proton Energy Systems, compañía dedicada a equipos electrolíticos, explica los argumentos económicos. El carbón, a los precios actuales —que probablemente se mantendrá más o menos en los próximos años—, cuesta unos euros por millón de kilocalorías. La gasolina (al precio actual en Estados Unidos de alrededor de medio dólar por litro, que, aunque hoy parece alto, podría ser inferior al del futuro) vendría a costar más de setenta dólares por millón de kilocalorías. Conseguir que las calorías del carbón se transformen en calorías capaces de propulsar un coche viene a ser como convertir agua en vino. Schroeder llama a su dispositivo, que convierte el agua en hidrógeno, "máquina de arbitraje". ("Arbitraje" es el término que los banqueros o los agentes de bolsa utilizan para describir una operación de cambio de valores mercantiles en la que se busca la ganancia aprovechando la diferencia de precios entre unas plazas y otras.) Cualquier política razonable debe ser consciente de los enormes diferenciales (diferencias entre precios) que existen en el mercado de valores energéticos.

Transporte y distribución del hidrógeno

Las diferentes fuentes de energía, sin embargo, quizá no puedan intercambiarse tan fácilmente como el dinero en un arbitraje. Suministrar hidrógeno de forma cómoda y barata no resulta sencillo, al menos si ha de obtenerse de fuentes renovables como la energía solar, hidroeléctrica o eólica, aprovechables sólo en ciertas áreas geográficas.

El hidrógeno procedente de aerogeneradores, por ejemplo, es competitivo en Estados Unidos con la gasolina cuando la energía eólica cuesta tres centavos por kilowatthora. Esto sucede en lugares donde siempre sopla el viento y, por tanto, poco poblados (en los EE.UU., la franja que va desde Montana y ambas Dakotas hasta Texas). Esta energía eléctrica de origen eólico tendrá que recorrer un largo trayecto hasta los usuarios finales, produciéndose pérdidas inevitables en la red de distribución. Para superar las limitaciones de la transmisión eléctrica, tal vez podría construirse un "gasoducto" que transportara, en forma de hidrógeno, la energía eólica generada en Dakota hasta Chicago —el mercado automovilístico más próximo.

Sin embargo, tampoco ésta constituye una estrategia factible. Dada la baja densidad del hidrógeno, su distribución resulta mucho más difícil que la del gas natural, por ejem-

EMISIONES DE LOS VEHICULOS

LAS EMISIONES de gases de invernadero (dióxido de carbono o equivalentes) varían según los efectos combinados del funcionamiento del vehículo y la fuente del combustible. Los vehículos con pila de combustible no emiten por sí mismos gases de invernadero, pero la fabricación de dichas pilas puede provocar emisiones que en total superen las de los motores de explosión o combustión interna (CI). (El Departamento de Energía estadounidense calcula que el etanol derivado del maíz casi no produce emisiones contaminantes, puesto que el dióxido de carbono que se libera en la combustión del etanol es reabsorbido por la siguiente siembra.)

LUCY READING; *FUENTE:* DEPARTAMENTO DE ENERGIA

plo. El transporte de cualquier gas en grandes volúmenes exige comprimirlo; de lo contrario, el gasoducto habría de tener el mismo diámetro que el fuselaje de un avión. La compresión requiere un trabajo que sustrae aún más energía del proceso productivo total. Por si fuera poco, la manipulación del hidrógeno es más complicada que la de otros gases combustibles. Un reciente estudio llevado a cabo por Ulf Bossel, Baldur Eliasson y Gordon Taylor, señala que el hidrógeno comprimido a 790 atmósferas posee menos de un tercio de la energía del metano presente en el gas natural a esa misma presión.

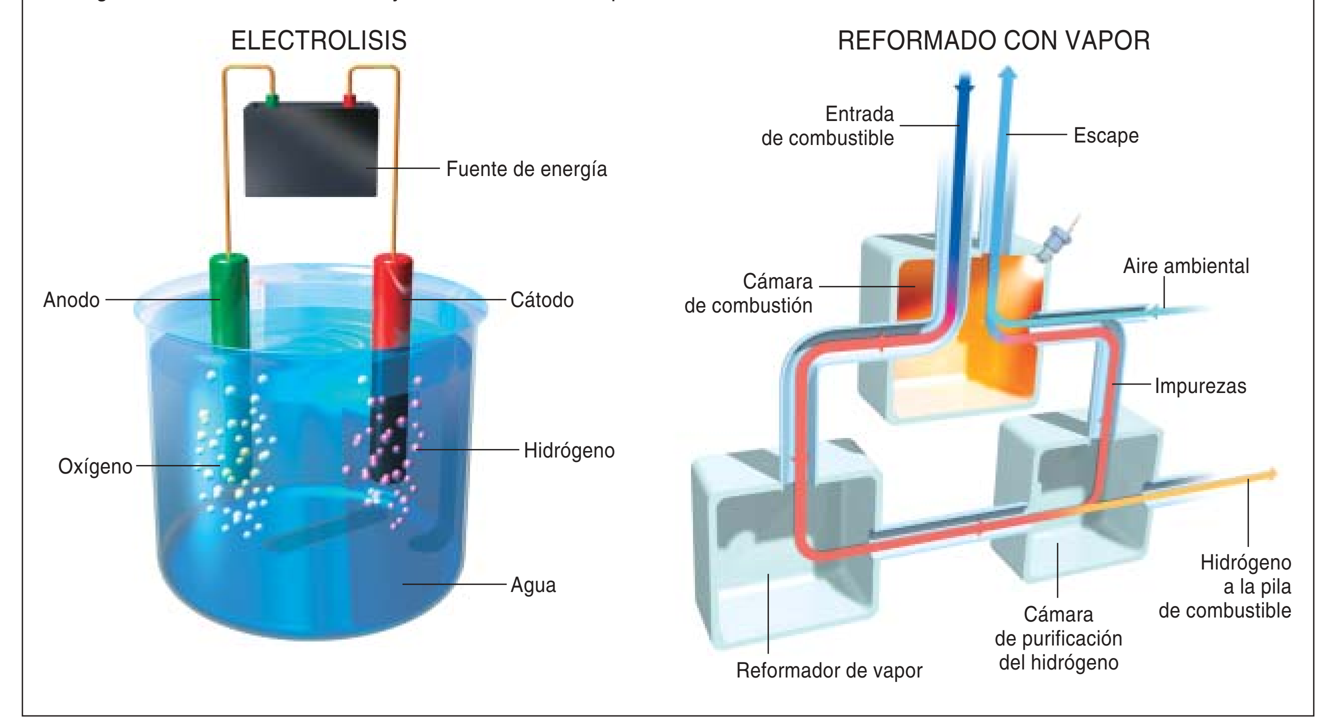
La misma investigación revela que un camión capaz de suministrar 2400 kilogramos de gas natural a un usuario sólo puede cargar 288 kilogramos de hidrógeno comprimido a la misma presión. En otras palabras, se necesitarían unos 15 camiones para cargar con el hidrógeno necesario para propulsar el mismo número de automóviles que abastece un solo camión de gasolina. Si se utilizara hidrógeno en estado líquido, tres camiones bastarían para conseguir el mismo objetivo; pero la licuefacción del hidrógeno exige un esfuerzo notablemente mayor. El transporte sería más fácil si se hiciera en forma de metanol (que se reformaría a bordo del vehículo para proporcionar hidrógeno); pero, una vez más, esta transformación adicional conlleva una penalización energética. Por tanto, resulta más conveniente utilizar el hidrógeno en su lugar de producción, que puede distar de los principales mercados de combustibles para la automoción.

Sea cual fuere la forma de distribuir el hidrógeno, ahí no acaban las dificultades que entraña la manipulación de este gas escurridizo. Otro de sus inconvenientes reside en su gran inflamabilidad. Todos los combustibles gaseosos arden a concentraciones comprendidas entre unos valores máximo y mínimo. El margen del hidrógeno es singularmente amplio: del 2 al 75 por ciento. El gas natural, en cambio, arde a concentraciones del 5 al 15 por ciento. Por tanto, si una fuga de gas natural resulta peligrosa, más lo será una fuga de hidrógeno, pues se incendiará dentro de un intervalo de concentraciones mucho mayor. La energía mínima necesaria para la ignición del hidrógeno también es mucho menor que para el gas natural.

Por si fuera poco, el hidrógeno arde de forma invisible. En un manual de seguridad publicado por la NASA, se recomienda vigilar los incendios de hidrógeno sosteniendo una escoba con el brazo extendido y comprobando si las pajas del extremo arden. Resulta pavoroso, en efecto, pues no se observa llama alguna. Michael D. Amiridis, de la Universidad de Carolina del Sur, entiende que un automóvil propulsado por pilas de combustible deberá operar con niveles de seguridad

OBTENCION DE HIDROGENO

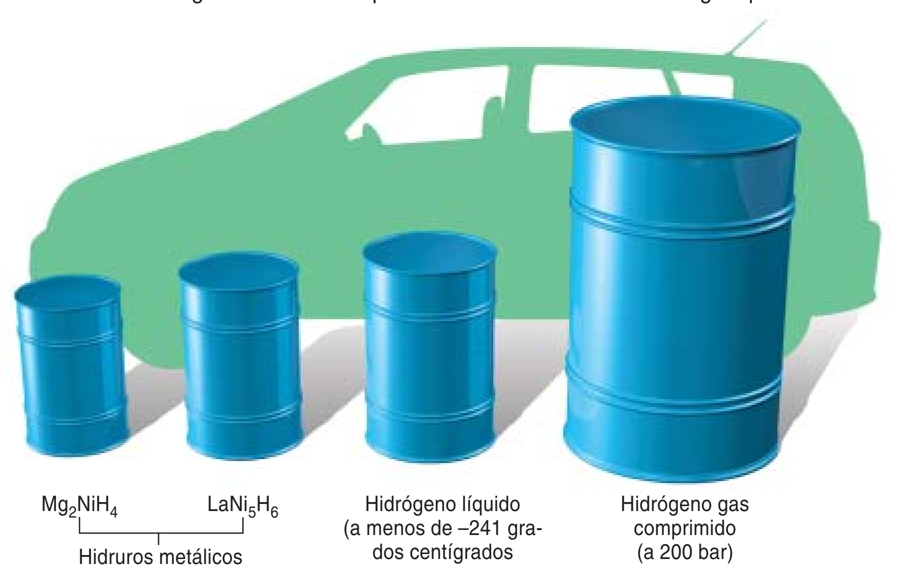
EN NUESTRO PLANETA, el hidrógeno puro (H_2) no se encuentra en la naturaleza. Se conocen dos métodos para obtenerlo. La electrólisis (*izquierda*) se sirve de corriente eléctrica para descomponer las moléculas de agua (H_2O). El cátodo (terminal negativo) atrae los átomos de hidrógeno; el ánodo (positivo), los de oxígeno. Ambos gases forman burbujas que salen a la atmósfera y pueden capturarse. En el reformado con vapor (*derecha*), un hidrocarburo como el metanol (CH_3OH) se vaporiza en una cámara de combustión caldeada. Luego, en el reformador de vapor, un catalizador media la reacción entre el combustible y el vapor de agua, que produce, entre otros componentes, hidrógeno. Por fin, éste se aísla y se conduce a una pila de combustible.



SLIM FILMS; *FUENTE:* IDATECH

LA MISMA CANTIDAD DE HIDROGENO EN DIFERENTES VOLUMENES

LLEVAR A BORDO DE UN AUTOMOVIL el gas más ligero del universo plantea un reto. Así lo evidencian las diferencias de volumen entre las distintas formas de almacenar los cuatro kilogramos de hidrógeno que se requieren para recorrer 400 kilómetros. (Cuatro litros de hidrógeno contienen aproximadamente la misma energía que 15 litros de gasolina; sin embargo, dado que el rendimiento de las pilas de combustible es aproximadamente el doble que el de los motores de explosión, los 4 kilogramos de hidrógeno proporcionarían un radio de acción igual al de 30 litros de gasolina.) Las soluciones actuales, incluso los depósitos de gas comprimido o hidrógeno líquido, ocupan demasiado espacio. Se está experimentando con hidruros metálicos u otros compuestos sólidos, que podrían desprender hidrógeno bajo demanda y recargarse después; pero también plantean inconvenientes: su elevado peso y el coste energético de las transformaciones químicas implicadas.



por lo menos equivalentes a los del coche actual. De hecho, en los primeros estadios del desarrollo de una cadena de suministro de hidrógeno combustible, gran parte del trabajo se ha centrado en la construcción de instrumentos de alarma capaces de detectar el gas hidrógeno con entera fiabilidad.

El papel del hidrógeno

Pese a los obstáculos técnicos y de infraestructura, tal vez nos estemos aproximando a una economía del hidrógeno. En tal caso, es muy posible que se parezca al mercado de los perfumes, en el que se comercializan cantidades tan minúsculas que los precios unitarios importan poco, a diferencia de lo que ocurre en el mercado de cereales, por ejemplo, donde una ligera variación del precio del kilogramo de producto se traduce en una gran diferencia en el precio total de la transacción, pues la mercancía se compra y se vende por toneladas. Es probable que se empiece por los teléfonos celulares y los ordenadores portátiles, a cuyos usuarios no les importará pagar alrededor de 10 euros por el kilowatt-hora procedente de pilas de combustible. La propia industria estima que este mismo año se empezarán a vender aparatos integrados en ordenadores portátiles. Eventualmente, el cambio podría llegar a nuestras viviendas, donde una potencia en torno a los cinco kilowatts bastaría para satisfacer una demanda de electricidad constante, que el carbón no siempre logra abastecer. En las carreteras, en cambio, resulta poco probable que lleguemos a ver muchas pilas de hidrógeno, dado que un automóvil exige una energía total de unos 50 kilowatts, pero no circula más de dos horas al día, por término medio: necesidades diametralmente opuestas a las prestaciones que ofrecen los dispositivos que, como la pila de combustible, tienen un coste de operación reducido y un coste por unidad de capacidad elevado. Aunque se habla mucho de las pilas de combustible como fuentes de energía alternativas para los coches, quizá sea ésta su aplicación menos comercial.

Para encontrar sustitutivos del petróleo en el sector del transporte, se pueden examinar varias alternativas antes de recurrir al hidrógeno. Una de ellas, el gas natural, con muy pocos detalles técnicos que resolver y amplia disponibilidad de suministro. Otra, la electricidad, para vehículos eléctricos. La alimentación por baterías ha tropezado con algunos obstáculos importantes, aunque podrían ser más fáciles de superar que los que presentan las pilas de combustible. En caso de necesidad, los vehículos pueden moverse con metanol obtenido del carbón; si los alemanes lo consiguieron en los años cuarenta, también está hoy a nuestro alcance.

En último término, si toda la sociedad apoyara el uso de aerogeneradores, células solares y otras fuentes de energía renovable, éstas reemplazarían grandes cantidades de combustibles fósiles que hoy se utilizan en la red de distribución eléctrica. Conjugando ese desarrollo con una política de conservación de los recursos naturales, dispondríamos de mucha más energía para aplicar al sector del transporte, precisamente el que más petróleo consume y el que más contribuye a la necesidad de potenciar alternativas como el hidrógeno.

Bibliografía complementaria

THE HYDROGEN ECONOMY: OPPORTUNITIES, COSTS, BARRIERS, AND R&D NEEDS. National Academies Press, 2004.

THE HYPE ABOUT HYDROGEN. Joseph J. Romm. Island Press, 2004.

SLIM FILMS; *FUENTE: NATURE*

Sida felino

Vacuna

En 1986 el equipo encabezado por Pedersen descubrió en el gato un virus similar al que produce la inmunodeficiencia humana (VIH), denominado virus de la inmunodeficiencia felina (VIF), que se asemeja al humano en sus características morfológicas, constitución genotípica y estructural y cuadro clínico del animal infectado. Por esta razón el VIF se utiliza como modelo de estudios patogénicos, terapéuticos y vacunales de la temida enfermedad humana (SIDA).

Al igual que el VIH, el VIF es un *Lentivirus* perteneciente a la familia *Retroviridae*. Parece, sin embargo, que le precedió en el tiempo, ya que los estudios retrospectivos demuestran la existencia de la infección por VIF en el gato desde mediados de 1968 en EE.UU. y en Japón, desde 1972 en Australia y desde 1974 en Francia. No significa ello que se iniciara en esas fechas, sino que no ha podido demostrarse su aparición con anterioridad.

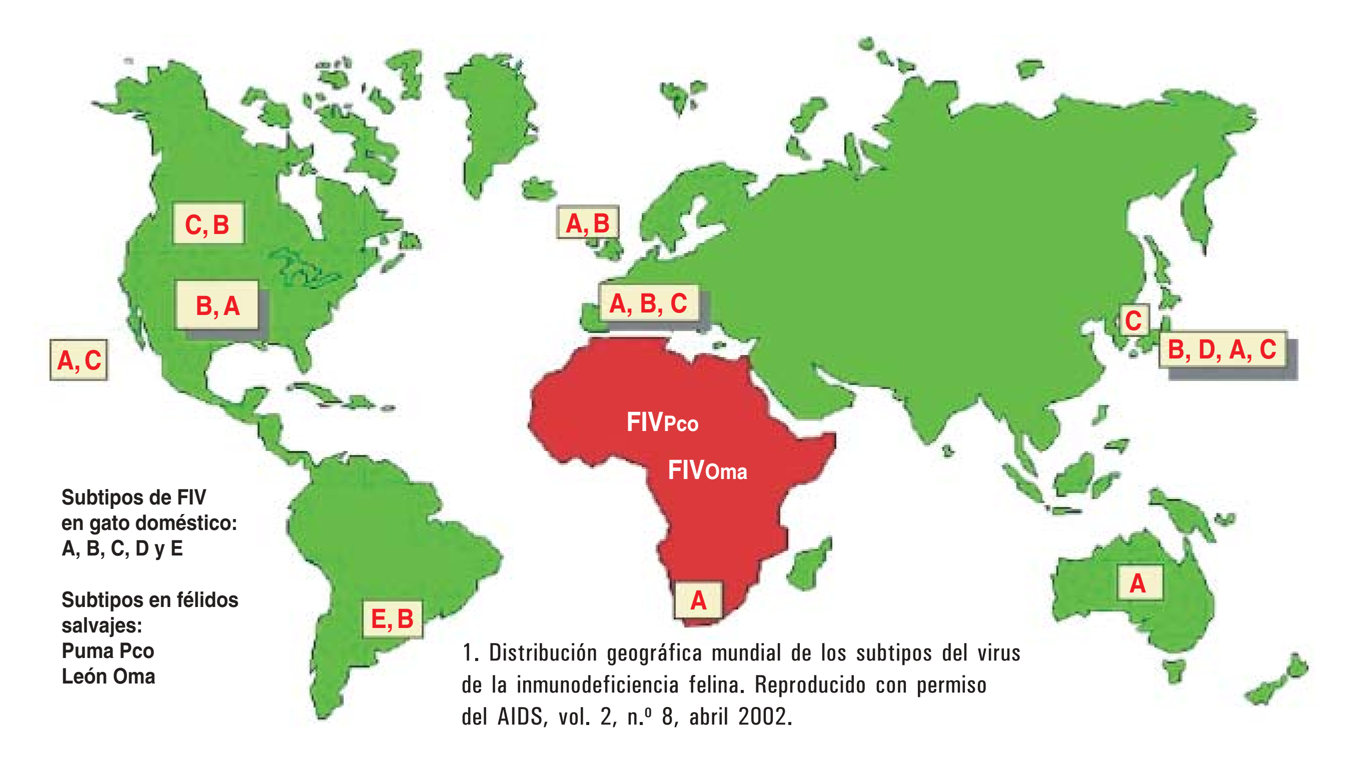
Existen algunas diferencias importantes con la infección humana. El virus de la inmunodeficiencia felina tiene una menor capacidad de elaborar intra e intersubtipos o de mutar en el gato infectado. En la especie felina se reconocen cinco subtipos (A, B, C, D y E). Los que gozan de una mayor distribución son A y B (*véase la figura 1*). Divergen también en la vía de transmisión; en el gato ocurre por mordedura. En cambio, la transmisión a través de mucosas, frecuente en los humanos, es rara entre los felinos.

La enfermedad felina persiste largo tiempo. Comienza con una fase breve (unas semanas a unos meses) que suele pasar inadvertida o muestra infecciones leves que responden al tratamiento. El VIF se multiplica en tejidos linfoides y glándulas salivales; el gato se convierte ya en infectante.

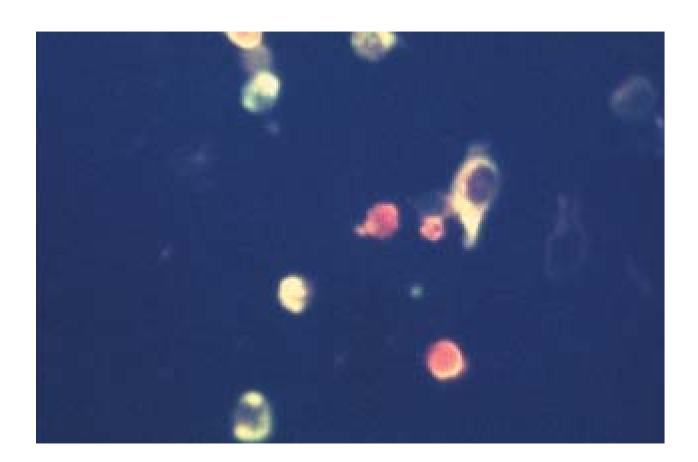
Se establece a continuación un período asintomático de 3 a 10 años. En el curso del mismo, el gato está sano en apariencia, pero la multiplicación vírica progresa. El virus puede ser aislado de linfocitos sanguíneos, suero, plasma, líquido cefalorraquídeo, semen, tejidos linfoides y saliva.

Al propio tiempo, y en paralelo, el gato muestra una elevada tasa de anticuerpos específicos frente al virus; no son suficientes para eliminar la infección, pues la protección debe establecerse también con una adecuada respuesta de células defensivas que destruyan las células infectadas. Durante esta fase se observa la pérdida progresiva de linfocitos T CD4+ y T CD8+, esenciales para la respuesta inmune frente a infecciones. Se trata de una fase extremadamente peligrosa para la diseminación de la infección, al no diferenciarse clínicamente los gatos portadores. Sólo podrá identificarse a los infectados mediante la detección de anticuerpos (*véase la figura 2*) específicos, el aislamiento y la reacción en cadena de la polimerasa.

La progresiva destrucción de células T desembocará en la fase fi-



1. Distribución geográfica mundial de los subtipos del virus de la inmunodeficiencia felina. Reproducido con permiso del AIDS, vol. 2, n.º 8, abril 2002.



2. Resultado positivo de la prueba de inmunofluorescencia indirecta utilizada para la detección de anticuerpos frente al virus de la inmunodeficiencia felina.



3. Infecciones respiratorias frecuentes que se presentan en la fase final de la evolución de la infección por VIF, o sida felino.

nal o sida felino. Lo mismo que en la especie humana, el gato sufrirá infecciones por agentes oportunistas, por virus y bacterias respiratorios (*véase la figura 3*), y digestivos, así como por infecciones emergentes (tuberculosis, infecciones por *Poxvirus*, hemobartonelosis, micosis sistémicas, neoplasias y otras). La muerte ocurrirá tras semanas o meses, sin que el tratamiento pueda detenerla.

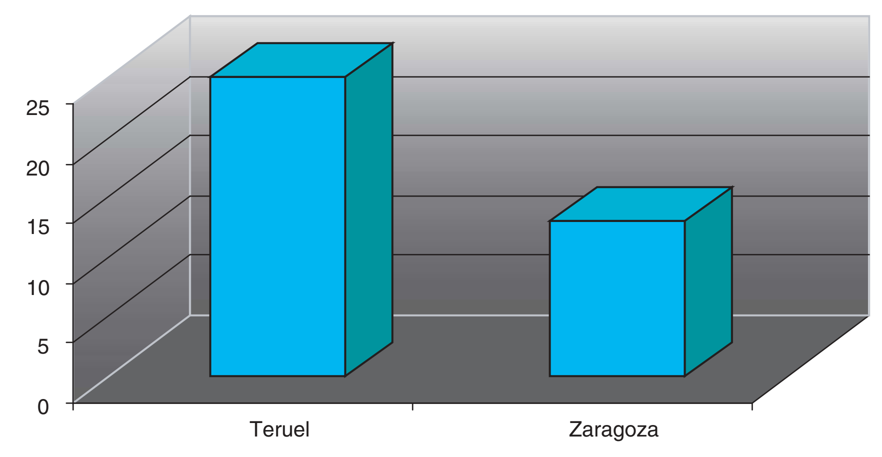
Al igual que el VIH, el VIF se halla ampliamente difundido. Se ha registrado su presencia en Estados Unidos, Canadá, Gran Bretaña, Europa continental, Japón, Australia, Nueva Zelanda y Africa del Norte y del Sur, con porcentajes de infección variables, desde el 0 % (Vietnam) al 30 % (Japón e Italia).

En el estudio realizado en la unidad de enfermedades infecciosas de la facultad de veterinaria de la Universidad de Zaragoza con gatos procedentes de Teruel y Zaragoza, observamos que un alto porcentaje de los gatos (18,5 %), estaban infectados. Los valores hallados en Teruel (25 %) superaban los observados en Zaragoza (13 %) (*véase la figura 4*). Esta diferencia significativa puede estar asociada al método de transmisión. La población felina de Teruel tiene más posibilidades de deambular por las calles y, en consecuencia, se halla expuesta a un número mayor de contactos agresivos: búsqueda de alimentos, peleas por las hembras, defensa del territorio, etcétera.

Los gatos mayores de 4 años presentaron una alta prevalencia (28 %), por el cúmulo de gatos infectados a lo largo del tiempo. Las razas comunes e híbridas se veían más afectadas (24 %), probablemente por provenir de, y convivir con, felinos callejeros. Mientras que los gatos de razas puras, generalmente bien cuidados, ofrecían unos valores de infección más bajos (11 %). Entre los gatos con síntomas persistentes (20 %) había también más portadores que en sanos (11 %). Se observó fluctuación de la prevalencia de infección según el año de estudio.

De nuestro trabajo deducimos la necesidad de tomar medidas preventivas para frenar la difusión de la inmunodeficiencia felina. Por resumir algunas: controlar las salidas al exterior de los gatos domésticos, castrar a los machos que viven en colectividad o están en riesgo de infección para reducir su agresividad, mantener a los animales bien alimentados y realizar visitas periódicas al veterinario, solicitando el diagnóstico de infección por VIF si el gato ha sido expuesto a animales infectados, vive en situación de riesgo, es mayor de 4 años o muestra síntomas persistentes de infección.

Una de las medidas más importantes es la vacunación. Se realiza ya en Estados Unidos. Puede iniciarse desde las 8 semanas de edad para gatos en riesgo. Se trata de una vacuna inocua, pues contiene las partículas víricas inactivadas. Se ha demostrado que protege frente a los subtipos heterólogos (B) y los homólogos contenidos en la vacuna (A y D). Está elaborada con cepas VIF aisladas de gatos con infección de larga duración no progre-



4. Diferencia de prevalencia de infección encontrada entre los gatos estudiados en Teruel y Zaragoza que han resultado significativas en el análisis estadístico.

siva; confiere inmunidad muy completa mediante anticuerpos neutralizantes, citoquinas, inmunidad mediada por células T coadyuvantes y células T citotóxicas. La vacuna supone no sólo un gran avance en la lucha frente a la infección felina, sino también una nueva esperanza en la lucha frente al sida humano.

M.ª del Carmen Simón Valencia
M.ª Victoria Martínez Alfonso
Unidad de Enfermedades Infecciosas y Epidemiología
Facultad de Veterinaria, Zaragoza

Neurociencias

El uso de antioxidantes en el tratamiento de las secuelas de la isquemia cerebral

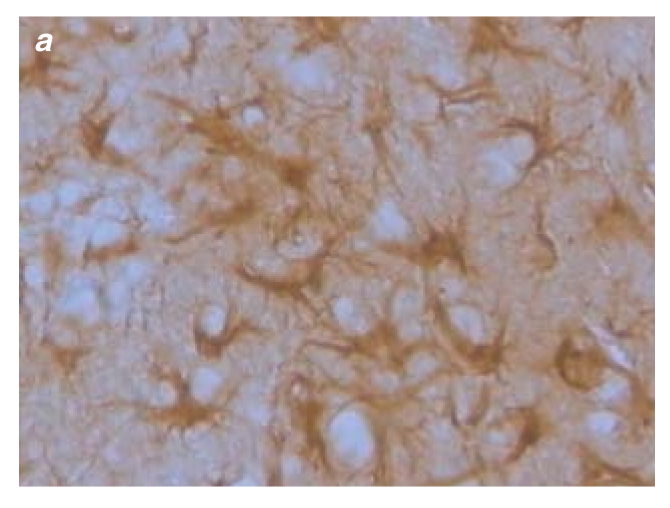
Durante las dos últimas décadas, la vitamina E (VE o α-tocoferol) y el ácido α-lipoico (AAL, ácido tióctico o lipoato) han centrado numerosos estudios farmacológicos. Por su carácter antioxidante, dichos compuestos podrían contrarrestar los efectos nocivos derivados de procesos oxidantes. En dichos procesos se genera un exceso de radicales libres (moléculas con uno o más electrones desapareados) que contribuyen al envejecimiento e intervienen en enfermedades degenerativas como la neuropatía diabética, el Alzheimer, el leucoma y la artritis reumatoide, entre otras. El AAL y la VE se presentan como potenciales agentes terapéuticos. Entre sus propiedades destacan su amplia disponibilidad en la dieta de mamíferos, buena tolerancia por vía oral, bajo costo y toxicidad mínima o casi nula.

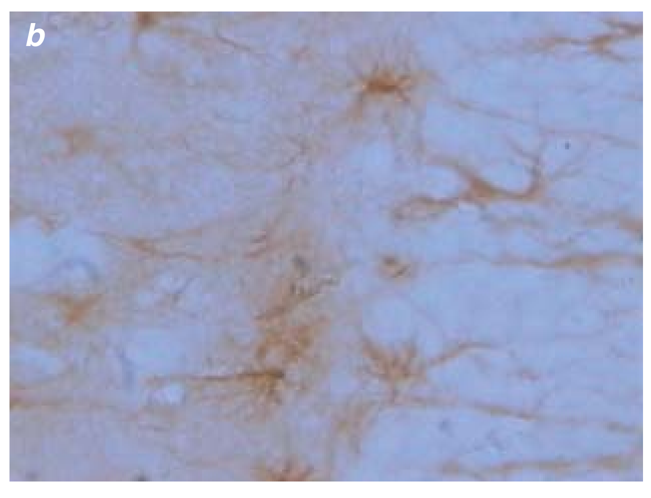
Desde hace más de un decenio se ha venido observando que el poder antioxidante del AAL depende de la concentración de VE. La interacción de ambas sustancias resulta claramente sinérgica. Estudios *in vitro* y experimentos de isquemia-reperfusión cardiaca (disminución-recuperación del flujo sanguíneo) demostraron que, en ausencia de VE, el AAL no inhibe la oxidación autocatalítica que degrada los fosfolípidos de la membrana celular (lipoperoxidación). Posteriormente se comprobó *in vivo* que esta mezcla de antioxidantes reducía la lipoperoxidación, mejoraba la actividad cardiaca, preservaba la contractilidad y disminuía la incidencia de arritmias cardiacas postinfarto. Así pues, la combinación de AE y AAL resulta mucho más eficaz que el consumo de sólo una de las dos sustancias.

Recientemente demostramos que la administración conjunta de AAL y VE durante 30 días previos a una isquemia cerebral facilita la posterior recuperación del tejido lesionado: acelera la recuperación funcional motora, reduce el tamaño del infarto (al favorecer la supervivencia de las neuronas en la zona dañada por falta de riego sanguíneo), disminuye la proliferación de astrocitos (células del sistema nervioso central que se tornan reactivas y forman la cicatriz glial) así como la respuesta inmune cerebral; también estimula la expresión de proteínas relacionadas con la remodelación nerviosa (sinaptofisina y GAP-43) y reduce los niveles de lipoperóxidos en sangre (fosfolípidos que han sufrido una degradación oxidativa inducida por radicales libres y, por tanto, constituyen una medida del estrés oxidativo).

También estudiamos el efecto de la mezcla antioxidante cuando ésta se administra inmediatamente después del episodio isquémico. Aunque no reduce significativamente el tamaño del infarto, sí produce sobreexpresión de las proteínas relacionadas con remodelación nerviosa. También disminuye la reactividad de los astrocitos (cicatriz glial), de forma comparable a la observada en el tratamiento antioxidante previo a la isquemia.

La estimulación de la remodelación neuronal e inhibición de la cicatriz glial se consideran eventos benéficos durante la reorganización morfológica y funcional que tiene lugar tras una lesión del tejido nervioso. Se ha sugerido que los efectos antioxidantes neuroprotectores de la combinación de VE y AAL favorecen dicha remodelación ner-





Microfotografía de astrocitos de hipocampo de pacientes isquémicos (40 aumentos). Se observa que la reactividad astrocitaria resulta mayor en el grupo control (*a*) que en el grupo tratado con la mezcla antioxidante de ácido lipoico y vitamina E (*b*).

viosa, en ausencia del obstáculo físico que representa la cicatriz glial. Estos resultados sustentan la posibilidad clínica de utilizar esta mezcla antioxidante para disminuir los efectos devastadores de un ataque cerebral o acelerar la rehabilitación de los pacientes, que en ocasiones se prolonga de 9 a 12 meses aun sin llegar a alcanzar una recuperación completa.

Pese a lo promisorio de esta posibilidad, no resulta evidente que la sobreexpresión de sinaptofisina (indicativa de la remodelación de sinapsis o sinaptogénesis) se traduzca en un aumento de la eficiencia del tejido afectado. De forma semejante, no podemos afirmar que el incremento de la proteína GAP-43, que se produce durante el rebrote de los axones (*sprouting*), culmine en conexiones funcionales. A este respecto sería necesario realizar estudios adicionales ultraestructurales de la zona lesionada, a mayores tiempos de recuperación.

En síntesis, la administración oral de la mezcla antioxidante de AAL y VE inhibe los efectos oxidantes de una isquemia cerebral y facilita la recuperación nerviosa (sin que necesariamente su consumo preventivo evite el infarto). Considerando la baja toxicidad y ausencia de efectos secundarios, creemos conveniente que la población con alto riesgo de sufrir eventos isquémicos —diabéticos, cardiópatas, hipertensos, sedentarios y fumadores, entre otros— ingiera dietas enriquecidas con estos dos antioxidantes biológicos. Así, de suceder un evento vascular-cerebral, disminuirá el tamaño del infarto, se inhibirá la cicatrización glial y se favorecerá la sobreexpresión de proteínas vinculadas a la remodelación nerviosa. Todo ello conducente a una mejor recuperación de las funciones dañadas.

J. García Estrada, R. E. González Castañeda y A. Navarro Ruiz
Centro de Investigación Biomédica de Occidente, Guadalajara, México

O. González Pérez, R. Ramos Zúñiga y S. Luquín
Depto. de Neurociencias, Centro Universitario de Ciencias de la Salud, Guadalajara, México

P. Garzón de la Mora
Instituto de Enfermedades Cronicodegenerativas, Guadalajara, México

Sondeo ionosférico y GPS

Combinación de datos de distinta naturaleza

La distribución variable en el tiempo y el espacio de los electrones libres de la ionosfera es de gran importancia para diversos ámbitos de la ciencia y la técnica.

La ionosfera es un plasma situado en la parte superior de la atmósfera. Se extiende, aproximadamente, desde los 60 hasta los 1000 kilómetros de altura con respecto a la superficie terrestre. La genera sobre todo la ionización solar; también interviene la precipitación de partículas tras las tormentas solares. La presencia de electrones libres en ella provoca en las señales electromagnéticas que la atraviesan un retraso, función de la frecuencia de las mismas (es decir, la ionosfera es un *medio dispersivo*).

Puede conocerse la distribución electrónica en la ionosfera gracias a los datos de los receptores del sistema de posicionamiento global (GPS). Para conseguir un sondeo más completo, se pueden complementar estos datos con observaciones efectuadas con ionosondas, sistemas emisores-receptores de pulsos electromagnéticos, radicados en tierra, que ofrecen información de la distribución vertical de electrones en la ionosfera.

El sistema GPS consta de una constelación de 24 satélites distribuidos en 6 planos orbitales, situados a unos 20.200 km por encima de la superficie terrestre. Cada uno de dichos satélites transmite un conjunto de señales moduladas en 2 frecuencias (1,2276 gigahertz y 1,57542 GHz) perteneciente a la banda L (un intervalo del espectro electromagnético en la región de las microondas).

Aprovechando la naturaleza dispersiva de la ionosfera y que los satélites del GPS transmitan en dos frecuencias, es posible, mediante una combinación adecuada de las señales del GPS, conocer el efecto que induce la ionosfera, sumado a otras contribuciones debidas a los retardos instrumentales de los transmisores y receptores. Dicho efecto o retraso es proporcional al contenido electrónico total (o densidad electrónica integrada) por unidad de área que haya a lo largo de la línea de visión entre satélite y receptor. Esta cantidad se denomina STEC (siglas en inglés de Contenido Electrónico Total Oblicuo).

Mediante la observación del STEC y métodos tomográficos similares a los usados en medicina, se obtiene una descripción tridimensional de la distribución de densidades electrónicas ionosféricas. El sistema GPS actúa de este modo como un gran escáner ionosférico planetario.

Uno de los varios métodos usados para obtener una descripción en tres dimensiones de la ionosfera consiste en dividir ésta en un retículo de celdas dispuestas en longitud, latitud y altura. Las observaciones del retardo ionosférico obtenidas mediante los receptores de tierra GPS se procesan conjuntamente para obtener la densidad electrónica en cada una de las celdas.

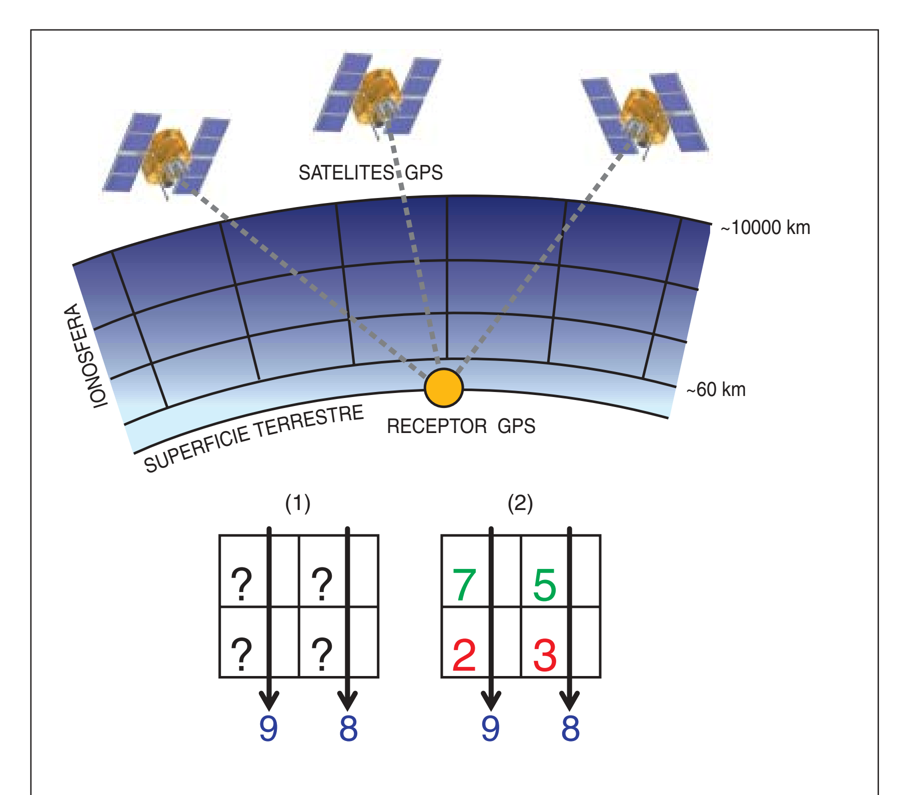
No obstante, uno de los problemas que plantea la tomografía mediante datos de tierra GPS es la poca resolución en altura inherente al planteamiento del método. Aunque existen organizaciones, como el International GPS Service (IGS), dotadas de vastas redes mundiales de receptores GPS que proporcionan una cobertura global para el sondeo ionosférico, los datos ligados a este tipo de receptores terrestres corresponden a observaciones básicamente verticales. Por eso, aunque se cuenta con una buena resolución horizontal, es virtualmente imposible descomponer la integral de las densidades electrónicas observadas en las diferentes capas verticales que conforman el retículo ionosférico. Se hace necesaria una fuente de información alternativa que permita separar la densidad integrada (observaciones GPS) para las distintas capas verticales (un ejemplo ilus-

trativo de esta necesidad se muestra en el recuadro).

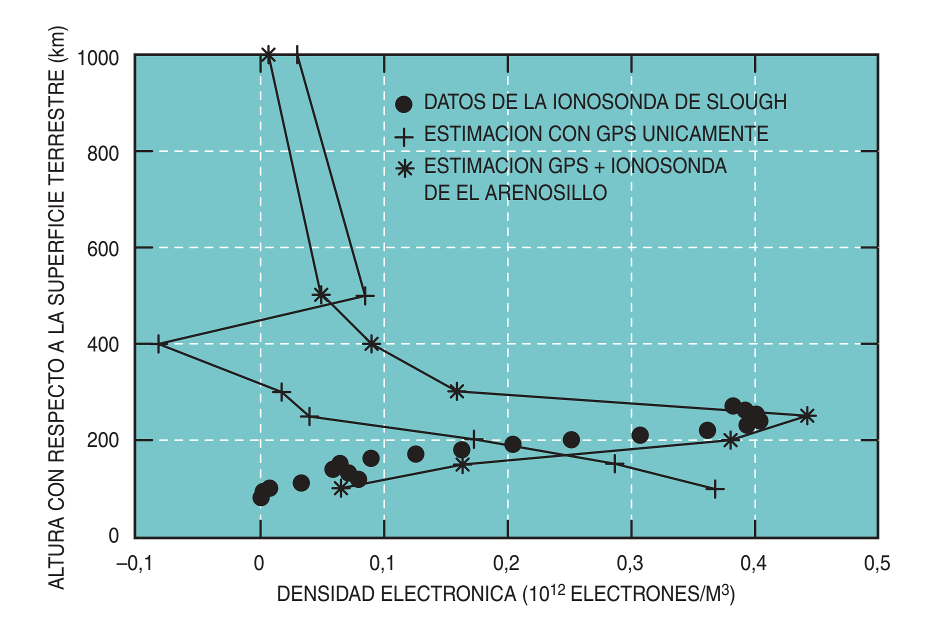
Las ionosondas ofrecen los datos necesarios para mejorar la limitada resolución vertical de los datos GPS terrestres. Estos instrumentos aprovechan la naturaleza dispersiva de la ionosfera para transmitir pulsos a distintas frecuencias. Sus rebotes dan información precisa de cómo se distribuyen verticalmente los electrones libres de la ionosfera. La cobertura de estos instrumentos, comparada con la de los receptores GPS, resulta limitada, por lo que ofrecen poca resolución horizontal.

Sin embargo, las propiedades complementarias de los datos de tierra obtenidos con el GPS y con ionosondas ofrecen la posibilidad de superar en cierta medida el problema de la poca resolución vertical. Para ello hay que combinar de manera adecuada los dos tipos de datos obtenidos de una red de estaciones de tierra GPS y varias ionosondas. De este modo, la falta de resolución vertical de los datos GPS se compensa con los datos de las ionosondas; la resolución vertical de éstas se expande en longitud/latitud gracias al uso de los datos GPS.

La resolución vertical de este método es limitada, ya que el número de variables se incrementa a medida que se añaden celdas verticales. Este problema se puede paliar reduciendo la retícula a una región; así se pueden conseguir resoluciones verticales de 50 km, en el mejor de los casos. De cara al futuro, con la mejora del método complementado gracias a una mayor potencia de calculo, será posible mejorar esta resolución y, por tanto, el sondeo ionosférico basado en GPS.



Las limitaciones a la hora de determinar la geometría vertical propias de las observaciones de tierra GPS plantean un problema en cuanto a resolución vertical de la estimación de la densidad electrónica de la ionosfera. El sencillo ejemplo de los recuadros de abajo presenta una analogía. Conociendo únicamente la suma de los recuadros de cada hilera vertical (*valores en azul, caso 1*) no es posible saber el contenido de cada recuadro. Ahora bien, si se provee información de alguno de los recuadros (*en verde, caso 2*), es posible deshacer la ambigüedad y determinar el valor de los recuadros restantes (*en rojo*). Los valores marcados en azul y en verde se corresponderían con los valores de STEC (el contenido electrónico total en la línea de visión entre receptor y satélite del GPS) obtenidos a partir de los datos de los receptores GPS y de las medidas de las ionosondas respectivamente.



Ejemplo de que valiéndose tan sólo del GPS no resulta posible obtener perfiles realistas de la densidad electrónica de la ionosfera. En el ejemplo, se muestran los valores del perfil de densidad proporcionados por la ionosonda de Slough (Reino Unido, 0,6 O 51,5 N), junto con las estimaciones, para esa localización, basados tan sólo en los datos GPS de 17 receptores distribuidos en Europa o complementados con los valores de otra ionosonda, situada en El Arenosillo (España, 6,7 O 37,1 N).

En geofísica, el conocimiento de los electrones libres de la ionosfera es de gran utilidad para la comprensión de esta misma y de los fenómenos físicos e irregularidades que en ella se desarrollan. En ámbitos más aplicados, las comunicaciones por ejemplo, su importancia estriba en que la ionosfera permite enlaces intercontinentales mediante el rebote de señales radioeléctricas en sus capas inferiores. En la navegación de satélites, aparecen cada vez más aplicaciones que exigen un error de posicionamiento pequeño. Para satisfacer dichos requerimientos y navegar con precisiones subdecimétricas, es necesaria una buena apreciación de la ionosfera.

MIQUEL GARCÍA FERNÁNDEZ
Dpto. Matemáticas.
Grupo de Astronomía y Geomática,
Universidad Politécnica
de Cataluña (UPC)
Barcelona

Calamares gigantes varados

Víctimas de exploraciones acústicas

Desde 1962 se han observado 47 registros de calamares gigantes (*Architeuthis dux*) en la costa occidental de Asturias. De éstos, casi el 75 % corresponden a animales vivos capturados por arrastreros que faenan entre 400 y 800 metros de profundidad en el extenso caladero de Carrandi, ubicado en las proximidades de tres cañones submarinos, principalmente para capturar bacaladilla (*Micromesistius poutassou*).

El ritmo natural de un registro anual de calamar gigante en esas aguas experimentó un significativo aumento en dos ocasiones. Entre el 13 de septiembre y el 23 de octubre de 2001 cinco calamares gigantes se hallaron varados cerca de los lugares donde los barcos *Barracuda* y *Nina Hay 502* trabajaron desde principios de septiembre hasta finales de octubre, utilizando dispositivos de cañones de aire comprimido para prospecciones geofísicas.

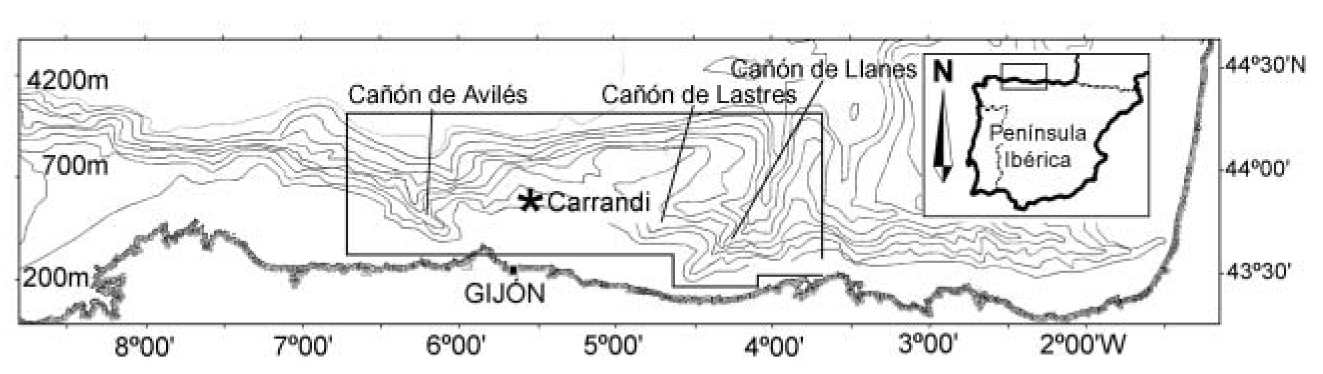
Dos años más tarde, entre el 13 y el 17 de septiembre de 2003, otros cuatro calamares gigantes vararon o fueron recolectados moribundos flotando en superficie en las proximidades de la costa. En esta ocasión, los cadáveres se recogieron en las cercanías de los lugares donde se desarrolló una campaña del proyecto internacional *Marconi*. Esta campaña de investigaciones geofísicas discurrió entre el 30 de agosto y el 18 de septiembre de 2003, y en ella se utilizaron también dispositivos con 10 cañones de aire comprimido para producir ondas acústicas de baja frecuencia (inferiores a los 100 hertz) y de alta intensidad (200 decibelios por cañón).

Todos los ejemplares aparecidos en ambos períodos pudieron ser conservados, seis se congelaron a –20 ºC inmediatamente después de su recolección y el resto fijados en formol al 10 % en agua de mar. Se realizaron las necropsias de estos nueve ejemplares de *A. dux*. Ocho de ellos resultaron ser hembras inmaduras o en maduración de entre 67 y 140 kg de peso, mientras el noveno era un macho maduro de 66 kg.

Siete de estos ejemplares están conservados en el Aula del Mar de CEPESMA en Luarca, otro en la sede de la Sociedad Gallega de Historia Natural en Ferrol, y el restante en el Museo del Mar que la Dirección General de Pesca del Principado de Asturias tiene en Gijón.

La hembra más grande (140 kg), que varó en la playa de Colunga el 13 de septiembre de 2003, tenía un manto cuyo grosor era de 35 mm y mostraba una banda de 43 cm de ancho en la zona central del manto con graves lesiones tisulares visibles. Las túnicas externa e interna de colágeno que envuelven la musculatura del manto seguían intactas, mientras que las fibras musculares y radiales que hay entre ambas túnicas aparecían destrozadas y convertidas en pequeños trozos. El grosor de la zona afectada era de 20 mm y mostraba intacto el reticulado romboidal de fibras de colágeno que hay entre las capas musculares del manto de estas y otras especies de calamares. Por otra parte, se observó que todos los trozos de fibras musculares estaban embebidos en un líquido cuya concentración de amonio era elevada.

Además de la destrucción diferencial de las paredes del manto, esta hembra mostraba la mayoría de los órganos internos prácticamente deshechos, formando un amasijo de tejidos dentro de la cavidad paleal: los órganos del tracto digestivo tenían varias largas y profundas llagas en distintas zonas,



1. Mapa donde se indican los tres cañones que se adentran en la plataforma de la costa asturiana y el centro geográfico del caladero de pesca de Carrandi.



2. Macho de *Architeuthis dux* de 66 kg varado en la costa oriental de Asturias.

especialmente en el ciego; la glándula digestiva estaba reventada, y su contenido, un fluido pardusco de consistencia oleosa, esparcido sobre el resto de la masa informe de tejidos; la capa externa del epitelio en columna de ambos corazones branquiales sufría desgarros en varias zonas, dando la impresión de haber explotado; las branquias, muy magulladas, mostraban varios capilares rotos; el ovario estaba roto en varios pedazos y numerosos ovocitos desprendidos se habían esparcido por toda la cavidad del manto. Aunque se buscaron, no se observaron signos de congestión vascular o microhemorragias. Ello podría deberse al relativamente pequeño volumen sanguíneo (aproximadamente un 5 % del peso corporal) y a la dificultad de observar hemorragias en animales cuya sangre carece de células y donde el pigmento respiratorio —la hemocianina— es de color azul claro.

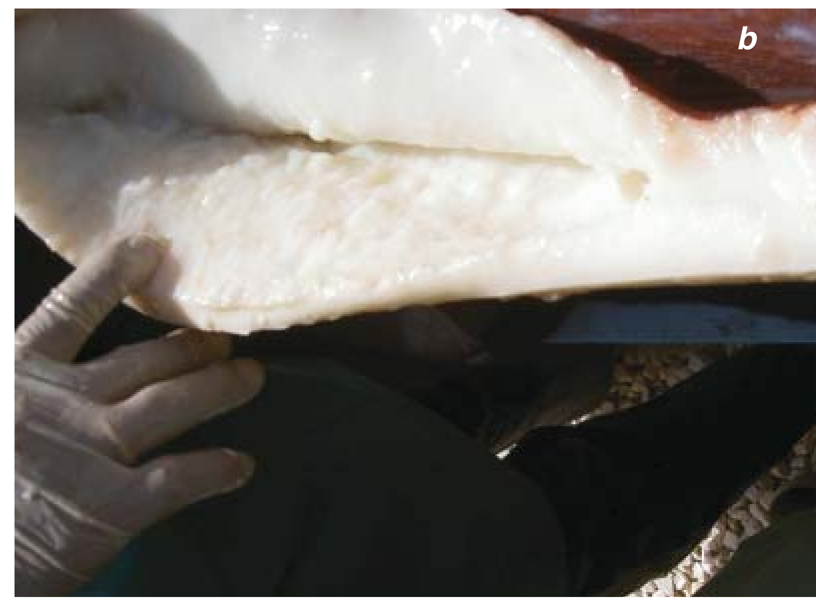
Se observó, además, que esa hembra imponente presentaba daños importantes en su sistema receptor del equilibrio o estatocistos. En los receptores de gravedad había algunas células foliculares receptoras de la mácula gravemente dañadas; dañadas, también, algunas células foliculares secundarias de las cristas del sistema de la recepción angular de ambos estatocistos.

Las lesiones observadas en las cinco de las seis hembras restantes y en el macho consistían principalmente en daños severos en sus estatocistos y en el sistema microvascular de sus branquias. Las concentraciones de Cd, Zn, Co, Cr y Ni en músculo del manto, glándula digestiva, corazón y branquias de tres ejemplares estuvieron dentro del rango de valores encontrados en otros cefalópodos pelágicos de zonas poco contaminadas.

En ninguno de los ejemplares estudiados se observó una excesiva carga parasitaria por metazoos. En ninguno se apreciaron signos o lesiones atribuibles a microorganismos, ni se aislaron bacterias patogénicas en los cultivos de tejidos obtenidos de tres cadáveres frescos.

Considerando el estado de madurez de los ejemplares, la mayoría inmaduros o madurando, quedó patente que su muerte no se produjo debido a la mortalidad natural posterior a su único ciclo reproductivo. No se observaron indicios de





3. Lesiones en las paredes del manto (*a*) y en los órganos internos de la hembra de mayor tamaño; (*b*) las capas de fibras musculares circulares y radiales de la pared del manto entre las túnicas externa e interna de colágeno de este ejemplar estaban destrozadas y convertidas en pequeños trozos.

otras posibles causas de muerte natural o de captura accidental.

Por el contrario, tanto las lesiones observadas como la concurrencia entre el sustancial incremento de los varamientos y la presencia de los barcos realizando prospecciones geofísicas con cañones de aire comprimido, sugieren que la muerte de estos animales pudo guardar relación con efectos subletales o letales producidos por el impacto de ondas acústicas.

El impacto de las técnicas acústicas marinas sobre el ambiente y los organismos marinos reviste particular interés, sobre todo en lo referente al efecto letal que las ondas de presión de diferente origen y naturaleza pueden causar en los oídos y en la activación del gas sobresaturado presente en la sangre de los mamíferos marinos y en sus células de diferentes tejidos, formando microburbujas que pueden provocar embolismos en órganos vitales. Este asunto ha dado lugar, entre otros, al trabajo de P. D. Jepson y colaboradores, publicado en *Nature* a finales de 2003, con réplica y contrarréplica muy recientes. No obstante, como se ha comprobado en numerosas reuniones internacionales, se conoce muy poco sobre el impacto de las técnicas acústicas marinas, tanto sobre los ecosistemas como sobre los organismos marinos, y mucho menos en el caso de los cefalópodos.

Hasta la fecha, además de los nueve ejemplares citados, hemos examinado diecisiete más, varados en diferentes partes de la costa española, y otros seis procedentes de Namibia. Esta es la primera vez que se ha observado el tipo de lesiones reseñadas en la mayoría de las hembras. Lesiones que, por su impacto diferencial en tejidos de distinta composición química y microestructura, así como la ausencia de llagas o heridas externas en el manto, no pueden atribuirse al varamiento o la captura.

Por otra parte, la presencia de daños en los receptores de ondas del sistema del equilibrio conduce a sugerir el siguiente mecanismo que explicaría su muerte: esas lesiones podrían haber provocado una importante desorientación en los animales, así como una disfunción en la información sensorial que permite regular su posición general y la de algunas partes de su cuerpo, en particular ante los giros. Además de esa desorientación, las ondas acústicas podrían también haberlos aturdido, debido a que son bastante sensibles a vibraciones de baja frecuencia.



4. Hembra inmadura de 70 kg de *A. dux* capturada por un pesquero en la zona más occidental del caladero de Carrandi en octubre de 2002.

Desorientados y aturdidos, estos cefalópodos moderadamente activos y con flotabilidad neutra podrían haber ascendido hacia la superficie, pasando de aguas profundas, cuya temperatura varía entre 10 y 12,5 ºC, a capas de aguas más superficiales, donde la temperatura puede superar los 15 ºC en la época del año en que se produjeron los acontecimientos. Ole Brix demostró que la capacidad de transporte de la hemocianina disuelta en la sangre de los calamares gigantes disminuía cuatro veces cuando la temperatura pasa de 6,4 a 15 ºC. Todo ello hace bastante plausible pensar que nuestros calamares podrían haber muerto asfixiados.

Puesto que los calamares gigantes flotan después de muertos y son suficientemente grandes como para que sus cadáveres tarden en ser devorados por los carroñeros, su presencia podría ser un indicador del impacto que ejercen las ondas de presión acústica en el ambiente y en la fauna marina de profundidad. Por otra parte, los cadáveres hallados en Asturias quizá sólo sean la punta de un iceberg que esté señalando efectos importantes sobre otras especies marinas, algunas de ellas de importancia comercial.

Es indudable que, para conocer a fondo los efectos de los pulsos acústicos de diferente intensidad y frecuencia en los organismos marinos de profundidad, es necesario intensificar los estudios físicos, fisiológicos y etológicos. Sin embargo, ante la ausencia de suficientes conocimientos, es recomendable adoptar una actitud precautoria. No se trata tanto de prohibir este tipo de técnicas, cuya utilidad es grande, cuanto de que todos los agentes implicados acepten y afronten el problema. Consideramos que se debería ser muy cauto a la hora de utilizar una técnica acústica hasta que su impacto en los ecosistemas y organismos se conozca suficientemente.

A. Guerra, A. F. González,
F. Rocha y J. Gracia
ECOBIOMAR.
Instituto de Investigaciones Marinas,
CSIC, Vigo

M. Vecchione
Museo Nacional de Historia Natural,
Institución Smithsonian,
Washington

DE CERCA

Jordi Corbera

Diversidad morfológica



1. *Pseudocuma ciliatum* (*arriba*). Cumáceo que medra en los fondos arenosos superficiales. Su cutícula escamosa y los pliegues oblicuos laterales le permiten penetrar rápidamente en el sedimento y así esconderse de sus depredadores. En el margen interno de su apéndice caudal (*derecha*) posee multitud de sedas con las que "cepilla" su propio caparazón para mantener la estructura escamosa en óptimas condiciones.

Los límites entre dos ecosistemas, los ecotonos, constituyen áreas de un gran interés debido a las relaciones que se establecen entre los organismos de los respectivos ambientes así como a la notable diversidad generada por la coexistencia de especies de ambos ecosistemas.

Lo observamos en el ecotono colindante entre los sedimentos marinos y el agua que los cubre. En los sedimentos, donde pululan organismos de dispar índole, abundan los gusanos poliquetos y los moluscos. Al otro lado de la frontera, sólo a unos pocos centímetros por encima del sedimento, las condiciones de vida resultan totalmente diferentes. Aquí domina una comunidad formada principalmente por pequeños crustáceos. Su capacidad natatoria les permite escapar de sus depredadores; además, muchos de ellos penetran hábilmente en el sedimento, lo que aumenta sus posibilidades de supervivencia. Entre los organismos mejor adaptados a este ecotono se encuentran los cumáceos, un grupo muy poco estudiado pese a haberse descrito casi 1400 especies. Presentan una morfología característica: caparazón globuloso y abdomen largo, cilíndrico y flexible, que recuerda las larvas de los mosquitos. Pero, la adaptación más singular se manifiesta en la cutícula, que, en respuesta a múltiples estrategias de defensa, ofrece una notable variedad de estructuras.

A diferencia de lo que ocurre en los peces, en el caparazón de los cumáceos las escamas se dirigen hacia adelante, mientras que en el abdomen lo hacen hacia atrás, para ayudar así la penetración en el sedimento. Estos crustáceos se entierran desde el centro del cuerpo (donde se sitúan las patas ambulatorias), levantando hacia arriba la cabeza y el extremo del abdomen, en forma de "uve". A menudo, en las especies que medran en fondos arenosos la cutícula escamosa se refuerza con una o varias carenas oblicuas que facilitan la penetración en sedimentos más gruesos. Algunas de las que viven a grandes profundidades o sobre sedimentos mucho más finos presentan, en cambio, una estructura menos marcada, lisa o incluso similar a la de las pelotas de golf, cubiertas por diminutos hoyuelos que minimizan el rozamiento. En otros casos, un caparazón aplanado les permite, igual que a lenguados y platijas, enterrarse con rapidez. Incluso existen especies que han optado por otro sistema de defensa: en vez de esconderse, han desarrollado fuertes espinas sobre su caparazón para disuadir a los depredadores.

FOTOGRAFIAS DE MICROSCOPIO ELECTRONICO COLOREADAS DIGITALMENTE Y REALIZADAS EN LOS *SERVICIOS CIENTIFICO-TECNICOS* DE LA UNIVERSIDAD DE BARCELONA



2. *Cumopsis goodsir*. Cutícula con estructura escamosa (orientada hacia adelante) y sedas sensoriales protegidas en pequeñas depresiones, que resultan esenciales para especies con una visión poco desarrollada.



3. *Cumella limicola*. Cumáceo que vive en el litoral. Sus ojos bien desarrollados captan los cambios de intensidad de la luz. Al atardecer empieza su migración vertical hacia la superficie.



4. *Platysympus typicus*. Cumáceo que habita en los fondos fangosos batiales del Atlántico occidental y del Mediterráneo. Su forma aplanada le permite un entierro frontal.

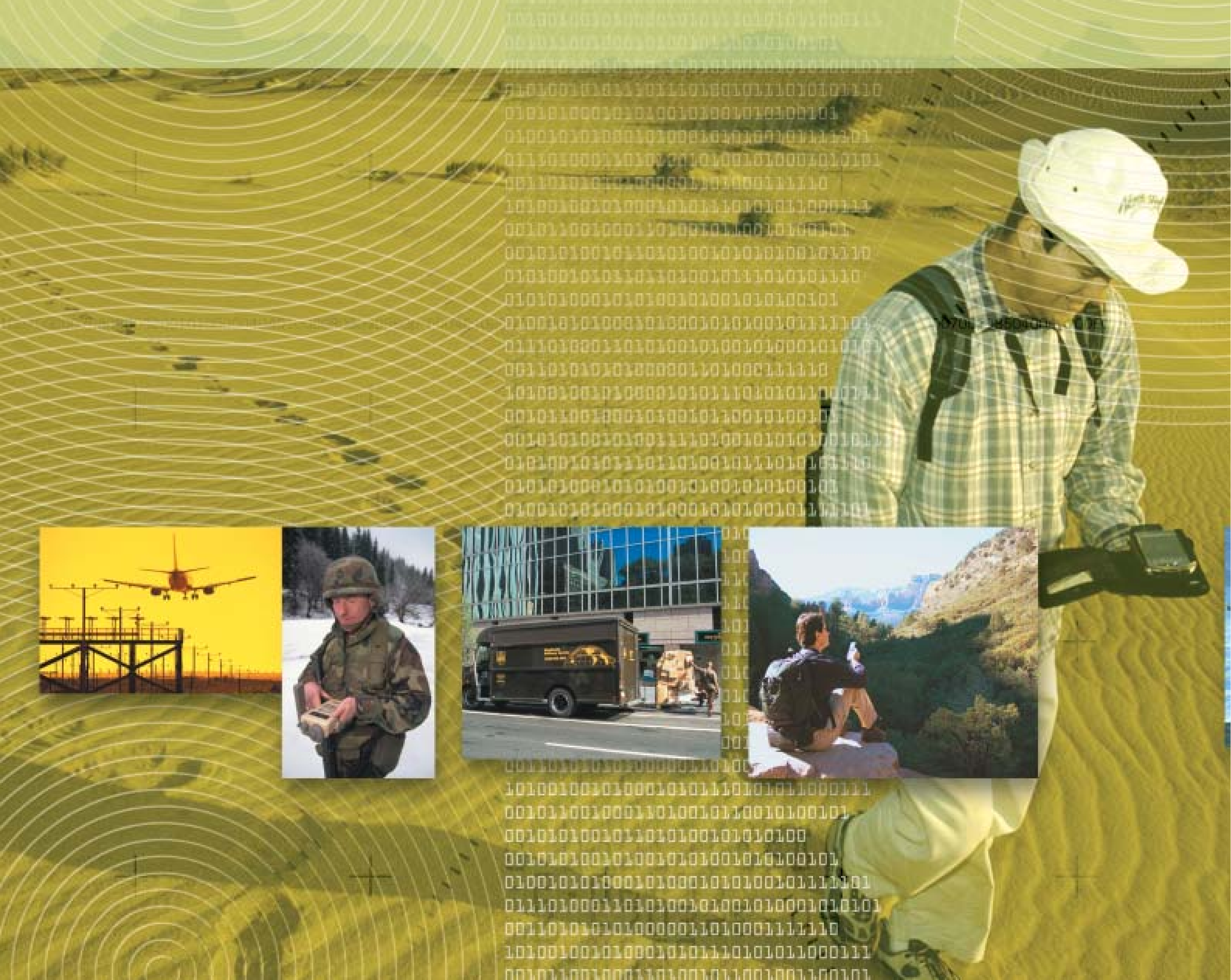


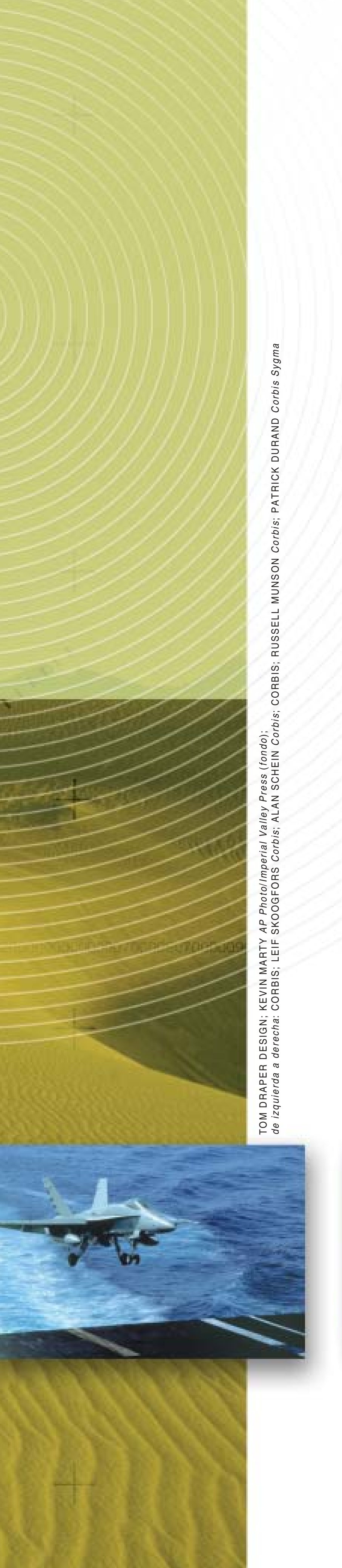
5. *Procampylaspis bonnieri*. La estructura reticular de su cutícula retiene una fina capa de sedimento. Además, los machos como el de la fotografía están provistos de fuertes espinas que disuaden a los depredadores.

La reestructuración del GPS

El sistema de localización global ya presta servicio a más de 30 millones de usuarios, sean excursionistas o pilotos que deben aterrizar con visibilidad nula.
Una mirada a lo que va a venir

Per Enge





TOM DRAPER DESIGN; KEVIN MARTY *AP Photo/Imperial Valley Press (fondo);*
de izquierda a derecha: CORBIS; LEIF SKOOGFORS *Corbis;* ALAN SCHEIN *Corbis;* CORBIS; RUSSELL MUNSON *Corbis;* PATRICK DURAND *Corbis Sygma*

Las prestaciones del Sistema de Localización Global, universalmente conocido por su sigla en inglés, GPS, van a mejorar de aquí a 10 años. El nuevo GPS no sólo contará con una seguridad y una fiabilidad mucho mayores que las posibles hoy en día, sino que localizará con mucha mayor precisión; su margen de error será de un metro. La mejora se funda en una serie de perfeccionamientos del sistema: la emisión de más y más potentes señales desde los satélites, la supervisión del funcionamiento, el cálculo de los errores máximos, la instalación de antenas inteligentes que seleccionan las direcciones de emisión o recepción de las señales, la integración del sistema en las redes de televisión y de telefonía móvil.

El sistema GPS de nueva generación permitirá un extenso repertorio de nuevas aplicaciones. Para la geolocalización, la cobertura abarcará desde las rutas de senderismo y las vías marítimas hasta los interiores y otras zonas con mala recepción, bajo unas ramas por ejemplo. Es probable que las empresas dedicadas al transporte, sea terrestre, marítimo o aéreo, la distribución de energía, las telecomunicaciones, la construcción, la minería, la topografía o la agricultura se beneficien con la ampliación de los servicios. Otro tanto les ocurrirá a las ciencias geográficas y geofísicas. No obstante, el mayor provecho será para el ejército, como pretendían los constructores del sistema original. Merced a la mayor fiabilidad del GPS reforzado, un avión podrá aterrizar con visibilidad nula, o un aparato militar posarse sin peligro en la cubierta de un portaaviones que se bambolea en la oscuridad. En años venideros, quizá hasta garantice la seguridad de quienes viajen por carreteras automatizadas.



EL GPS, cada vez más utilizado, adquirirá la fiabilidad suficiente para aplicaciones donde la seguridad es esencial, como la conducción automática de vehículos terrestres o aéreos.

El cielo es el límite

El sistema GPS entró en servicio en 1978 con el lanzamiento del primer satélite Navstar por el Departamento de Defensa estadounidense. Aunque los proyectistas del sistema confiaban en que tuviera aplicaciones en la vida civil, su intención en aquel entonces era proporcionar a unos 40.000 usuarios militares un sistema de navegación por tierra, mar o aire de alta precisión. Los civiles empezaron a sacar partido del sistema en los años ochenta, y al ir llegando la constelación orbital de satélites GPS al número mínimo de los 24 requeridos para prestar servicio ininterrumpido, lo que ocurrió a comienzos del decenio siguiente, no tardaron en multiplicarse las aplicaciones orientadas a grandes mercados.

Son en la actualidad unos 30 millones quienes regularmente se valen del GPS para determinar dónde se encuentran. Las unidades receptoras facilitan la conducción de vehículos terrestres, el pilotaje de barcos y aeronaves y el control de los coches de alquiler; sirven también para muchos usos recreativos. Todos los meses son expedidos al mercado mundial más de 200.000 receptores para uso civil, por valor, en el año 2003, de más de 3000 millones de euros. Esta cifra podría alcanzar los 10.000 millones a partir del año 2010, según un estudio reciente de Frost & Sullivan, una firma de investigación de mercados. En tales cifras no se han contado los ingresos correspondientes a la construcción, lanzamiento o control de satélites, ni tampoco los de las empresas relacionadas con el GPS, como las de gestión de flotas de vehículos de reparto. De acuerdo con el estudio, algo más de la mitad de las ventas de equipos corresponden a consumidores individuales, un 40 por ciento a empresas y el 8 por ciento restante a las fuerzas armadas.

Pero no son los Navstar estadounidenses los únicos satélites geolocalizadores que se hallan en órbita. Los satélites rusos GLONASS comparten ese mismo espacio físico y funcional; dentro de pocos años lo hará la constelación Galileo de satélites europeos. La Unión Soviética construyó el sistema GLONASS durante la Guerra Fría para competir con el ejército estadounidense. En estos últimos años GLONASS ha caído en desuso porque sus operadores no pueden costear el reabastecimiento de sus satélites. Se es-

pera que el sistema de la Comunidad Europea entre en servicio en el decenio en curso. El incentivo es el aumento explosivo previsto en el mercado de usuarios finales, conforme vayan siendo incorporados los receptores de GPS a los automóviles y a los teléfonos móviles. Europeos y rusos están convencidos de que para participar necesitan sistemas propios de satélites de navegación. Los directivos del GPS y de Galileo han establecido recientemente acuerdos sobre la interacción de ambos sistemas.

Cada vez que un receptor de GPS determina su posición sobre la superficie del planeta, efectúa una "trilateración" (algo parecido a una triangulación) de su distancia exacta a un mínimo de cuatro satélites GPS valiéndose de señales telemétricas emitidas por éstos (*véase el recuadro* "Así funciona el GPS"). En esencia, las señales de radio, que están codificadas de un modo especial, actúan a modo de metros invisibles que evalúan la distancia desde los satélites al receptor.

La precisión de un receptor manual que cueste ciento y pico euros sitúa al usuario dentro de un radio de unos 5 a 10 metros alrededor de su auténtica ubicación. Una unidad militar, de mayor precio, determina su posición con un margen de menos de cinco metros. Mediante observaciones simultáneas, en tándem, en las que el receptor cuenta con correcciones de error proporcionadas por otro receptor estacionario, instalado en un punto de coordenadas conocidas, es posible alcanzar precisiones de medio metro. Estos sistemas de funcionamiento en tándem reciben el nombre de GPS diferencial.

Los datos que llegan del espacio

Para apreciar adónde se encamina el GPS convendrá echar primero una mirada a su funcionamiento actual. La "precipitación" de datos transmitidos por los satélites GPS se parece a una tenue llovizna. Un satélite GPS radia señales con una potencia de 500 watt, equivalente a cinco bombillas de incandescencia. Tras viajar 20.000 km por el espacio, las señales de radiotelemetría llegan a la superficie terrestre con una densidad de potencia de 10^{-13} watt por metro cuadrado. Como referencia, la potencia de las señales de televisión que llegan a un receptor doméstico es mil millones de veces mayor.

Los satélites GPS rocían la Tierra con información de dos tipos. Uno de ellos, el mensaje de navegación, consiste en ristras de datos binarios que identifican la situación orbital del satélite y el instante en que se envió la transmisión. El segmento de control de GPS, que utiliza una red de receptores de GPS instalados en puntos de referencia, calcula en tierra esas coordenadas espaciotemporales. Estos valores son transmitidos al satélite mediante un radioenlace e incrustados en el mensaje de navegación para su posterior retransmisión a todos los usuarios.

El segundo tipo de información emitida por los satélites GPS consiste en un sistema de códigos de telemetría, compuestos por una secuencia de impulsos digitales que obedecen a una pauta peculiar y bien determinada. Las emisiones de este tipo no portan datos en el sentido habitual de la palabra. Están concebidas para la medición en el receptor del instante de llegada de la señal entrante, medición clave para la determinación precisa de la posición. Para resaltar la diferente naturaleza de estas señales de telemetría, se dice que estos "códigos de ruido pseudoaleatorio" (RPA) no constan de bits, sino de "chips".

Cada secuencia de códigos RPA viene a ser como la serie de notas de una pieza musical. Supongamos que el satélite y el receptor estuviesen tocando con exacta simultaneidad una misma canción. El usuario podría oír ambas versiones de la canción (vale decir, del código RPA), pero la versión del satélite se retrasaría por el tiempo que tardará el sonido en llegar desde la órbita hasta la superficie terrestre. Si el usuario cronometrase el desfase entre los instantes en que cada versión de la canción alcanza una determinada nota, podría determinar cuánto tarda el sonido en recorrer la distancia desde el espacio. Multiplicando el número de segundos resultante por la velocidad del sonido se podría calcular la distancia al satélite.

El GPS lleva a cabo un procedimiento parecido cuando un receptor supervisa el código RPA que está siendo emitido por un satélite. Al cotejar la secuencia telemétrica recibida (la secuencia de notas musicales) con una copia de la secuencia RPA peculiar de ese satélite, que se encuentra almacenada en el receptor, el dispositivo puede estimar el retraso en la llegada de la señal de telemetría. El receptor procede entonces a multiplicar el retraso cronometrado por la velocidad de la luz, y de este modo determina la distancia al satélite.

Así pues, los receptores evalúan la distancia mediante un "metro virtual" que cada satélite extiende hasta la superficie terrestre. Los códigos proporcionan marcas de distancia sobre la "cinta métrica herciana", mientras que el mensaje de navegación describe la posición del satélite, que viene a ser como el otro extremo de la cinta. Si la unidad de GPS estuviera dotada de un reloj perfecto, bastarían tres medidas de distancia para que el receptor pudiera resolver en tres dimensiones su posición, es decir, su latitud, longitud y alti-

Resumen/GPS reforzado

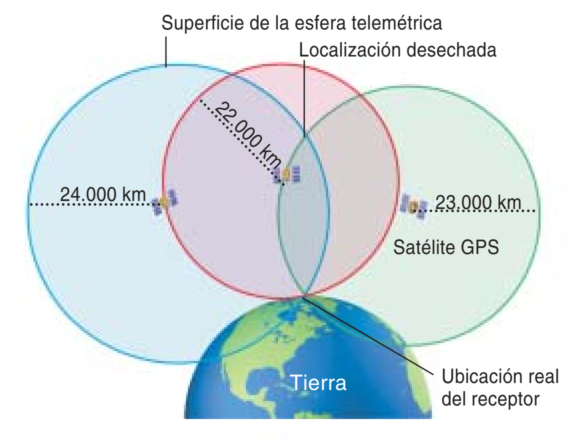
- Más de 30 millones de personas recurren al GPS (Sistema de Localización Global) de forma regular. Ese número no tardará en multiplicarse cuando se integren los receptores en más teléfonos y automóviles.
- Se obtendrá una geolocalización de precisión muy superior cuando estén disponibles las nuevas señales para usos civiles o militares. Se dispondrá de la primera de ellas cuando entre en órbita una nueva serie de satélites perfeccionados, en el año 2005. La segunda entrará en servicio algunos años después.
- Las máquinas de integridad del GPS garantizarán la fiabilidad del GPS proporcionando en tiempo real cotas válidas de error. Una gama de actividades técnicas, operativas e institucionales fortalecerán las emisiones de GPS frente a las interferencias de radiofrecuencia.

ASI FUNCIONA EL GPS

EL SISTEMA DE LOCALIZACION GLOBAL (GPS, Global Positioning System) es un sistema de navegación que cubre la superficie terrestre. Consta de dos docenas de satélites y de estaciones terrestres asociadas. Mediante un procedimiento similar a la triangulación, la trilateración, el GPS calcula las distancias existentes hasta, al menos, cuatro de los satélites. Son varios los factores que, combinados, permiten una geolocalización precisa.

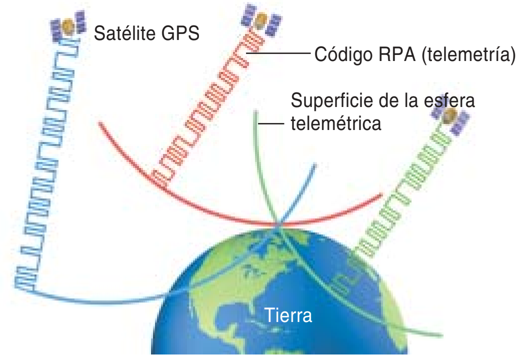
ESFERAS SECANTES

Supongamos que un receptor de GPS determina que la distancia a un satélite es de 22.000 kilómetros. Consecuentemente, el receptor debe encontrarse en la superficie de una esfera de 22.000 kilómetros de radio cuyo centro es el satélite. Supongamos que el receptor estima también que la distancia a otros dos satélites es de 23.000 y 24.000 kilómetros, respectivamente. El receptor ha de encontrarse, por necesidad, en la intersección de estas tres esferas telemétricas. Tres esferas de distinto centro no pueden cortarse en más de dos puntos. Sólo una de estas dos posiciones podrá estar lo suficientemente próxima a la superficie terrestre como para corresponder a la ubicación del receptor.



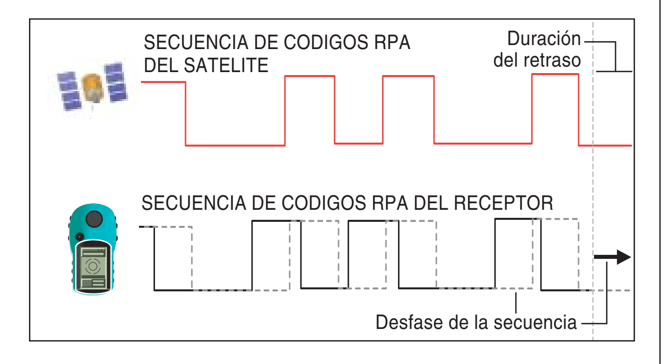
SEÑALES DE CRONOMETRAJE

La evaluación de la distancia a un satélite se logra cronometrando el tiempo que tarda una señal emitida por el satélite en llegar al receptor. El producto del tiempo de viaje por la velocidad de la señal —la velocidad de la luz, aunque para mayor precisión habrá que tener en cuenta su dependencia de las condiciones eléctricas de la atmósfera— proporciona la distancia recorrida. El problema es la medición del tiempo de viaje. Una compleja secuencia de datos binarios, el llamado ruido pseudoaleatorio (RPA), contribuye a realizar dicha tarea. Cada satélite tiene un código propio y peculiar, para evitar que el receptor confunda las señales.



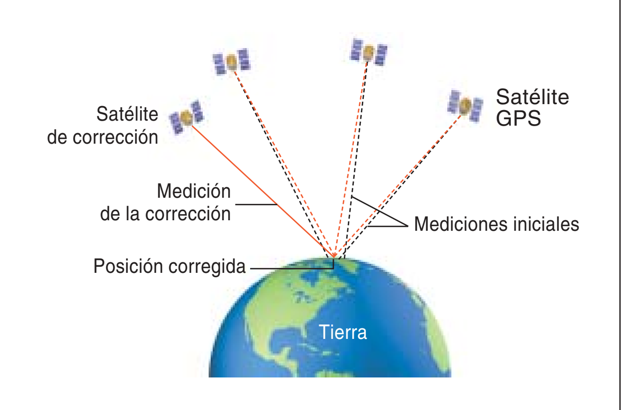
DESFASES EN LOS CODIGOS

Para comprender la forma en la que los códigos RPA permiten la medición de distancias, tomemos este símil: supongamos que tanto el satélite como el receptor empezasen a tocar una misma canción —el código RPA— al mismo tiempo. La versión emitida desde el espacio llegaría con cierto retraso con respecto a la canción de referencia del receptor. Midiendo cuánto tardó una determinada nota —un segmento del código RPA— en llegar desde el satélite, el sistema podría determinar el tiempo de viaje. Multiplíquese ese tiempo por la velocidad de la luz, y el resultado será la distancia al satélite.



SINCRONIZACION DE RELOJES

El tiempo se controla muy estrictamente a bordo de los satélites GPS: cada uno de ellos va dotado de un reloj atómico. En cambio, los receptores tienen que arreglárselas con un reloj de cuarzo de poco precio, cuya precisión es mucho menor. A causa de las deficiencias inherentes al cronometraje, las tres esferas de la telemetría no se intersecan exactamente (*líneas de puntos negras*). Para sincronizar los relojes situados en órbita con los ubicados en tierra, y compensar así el error de cronometraje, el GPS ha de efectuar una medida a un cuarto satélite. El valor obtenido determina un único coeficiente de corrección, que hará coincidir las tres mediciones de telemetría con la auténtica ubicación del receptor (*señalada en rojo*).



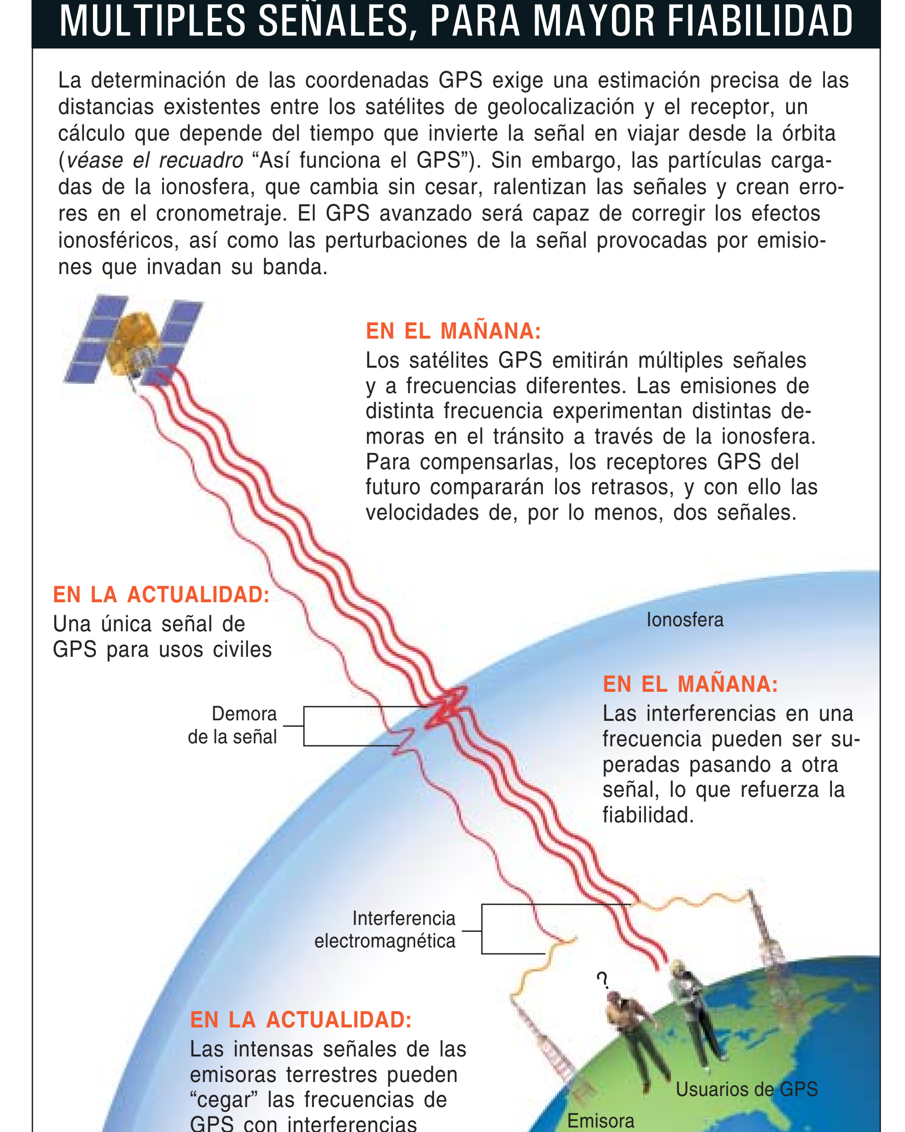
ALFRED T. KAMAJIAN

tud. Si los relojes fueran perfectos, bastaría una sola medida para situar al punto en la superficie de una esfera con centro en el satélite y radio prescrito. Dos mediciones de GPS restringen la posición del usuario a la intersección de dos esferas, mientras que tres lo sitúan en un único punto, definido por las tres esferas y la superficie de la Tierra. De este modo, el receptor habría de resolver un sistema de tres ecuaciones con tres incógnitas, la latitud, la longitud y la altitud.

Pero los relojes perfectos no existen, y por ello los receptores de GPS tienen necesidad de resolver una cuarta incógnita: la deriva entre el reloj interno del receptor, de bajo precio, y el tiempo de la red GPS, que está controlado con un error de menos de una milmillonésima de segundo mediante relojes atómicos; en cambio, el reloj del receptor podría sufrir un error de un segundo, o más, al día. El error de tiempo se convierte en un error de distancia al multiplicarlo por la velocidad de la luz. Esta deriva añade un número incógnito a la distancia estimada por cada satélite, lo que explica por qué las mediciones de longitud reciben el nombre de mediciones pseudotelemétricas. Felizmente, la deriva en tiempo es la misma con respecto a todos los satélites, por lo que una medición tomada de un cuarto satélite permite que el receptor resuelva las cuatro ecuaciones necesarias para la determinación de las cuatro incógnitas: longitud, latitud, altitud y tiempo.

Debido a que los usuarios pueden cambiar de posición rápidamente, los receptores de GPS actuales supervisan también la fluctuación en frecuencia de las señales entrantes atribuida al efecto Doppler, es decir, las variaciones de la longitud de onda de las señales recibidas debidas al movimiento. En el caso de que el usuario se esté alejando del satélite, parecerá que aumenta la longitud de onda de la señal; si el usuario se mueve hacia el satélite, la longitud de onda se acortará. Cada satélite viene a ser como un tren que pasa junto a una persona (el receptor). Cuando se acerca, el tono del silbato se vuelve más agudo; cuando se aleja, más grave. La supervisión de estas variaciones en la longitud de onda permite a esos equipos la estimación de la velocidad del usuario de forma directa y más precisa.

Resulta notable que los receptores de GPS consigan llevar a cabo la compleja tarea de la geolocalización sin transmitir señal alguna. Gracias a ello, los receptores que se instalarán en futuros teléfonos móviles serán muy económicos: no costará cada uno ni 5 euros.



MULTIPLES SEÑALES, PARA MAYOR FIABILIDAD

La determinación de las coordenadas GPS exige una estimación precisa de las distancias existentes entre los satélites de geolocalización y el receptor, un cálculo que depende del tiempo que invierte la señal en viajar desde la órbita (*véase el recuadro* "Así funciona el GPS"). Sin embargo, las partículas cargadas de la ionosfera, que cambia sin cesar, ralentizan las señales y crean errores en el cronometraje. El GPS avanzado será capaz de corregir los efectos ionosféricos, así como las perturbaciones de la señal provocadas por emisiones que invadan su banda.

ALFRED T. KAMAJIAN

A través de la ionosfera

Los equipos instalados a bordo de los satélites GPS emiten su información mediante ondas de radiofrecuencia (RF) estándar. La portadora de RF es la clásica sinusoide; su frecuencia indica el número de ciclos por segundo (compuestos cada uno por una cima y un valle). La técnica de GPS actual utiliza dos bandas de frecuencia, L1 y L2, situadas en la región de microondas del espectro electromagnético. A la banda L1 suele llamársela "señal civil", a pesar de que este recurso se comparte con las fuerzas armadas. Es accesible a todos y soporta la casi totalidad de las aplicaciones civiles actuales. La banda L2 es de uso casi exclusivamente militar. El público puede usarla, pero no se le permite conocer los códigos RPA militares. Esta ignorancia hace que la aplicación de la banda L2 para usos civiles sea frágil. Los receptores civiles, por ejemplo, tienen di-

ficultades en utilizar la señal L2 de satélites que se alcen poco sobre el horizonte o estén oscurecidos por obstáculos, aunque sean de pequeño tamaño, como los árboles. Además, los receptores L2 son caros, porque exigen técnicas especiales de procesamiento de la señal para tener acceso a las señales de esa banda cuando se carece de los códigos RPA.

Por las razones explicadas, la inmensa mayoría de las unidades de uso civil utilizan solamente la señal L1. Al hacerlo, consiguen normalmente una precisión de entre 5 y 10 metros, un margen de error debido en gran medida a las partículas cargadas de la ionosfera terrestre, que se extiende desde una altitud de unos 70 kilómetros sobre el suelo hasta unos 1300 kilómetros o más. Esta envoltura conductora frena la transmisión de las ondas de radio desde los satélites GPS, al igual que el agua de un vaso dobla, o refracta, la imagen de un lápiz inmerso en ella. Según las condiciones, puede demorar la llegada de una transmisión entre 1 y 10 metros, o más.

Con el fin de corregir estas desviaciones, algunos usuarios utilizan el llamado GPS-D, o GPS diferencial. Se vale de dos receptores GPS: una unidad móvil y otra, de referencia, emplazada en una ubicación conocida. El dispositivo de referencia transmite al receptor en movimiento las diferencias entre los valores de sus mediciones de distancia a los satélites y las distancias a que calcula que se encuentran. El receptor móvil utiliza tales datos para corregir la determinación de su situación. El funcionamiento del GPS diferencial es óptimo cuando el receptor móvil se mantiene no muy lejos del receptor de referencia. A distancias de menos de 100 kilómetros, los errores ionosféricos se cancelan casi totalmente, porque el haz de radio que va desde el satélite hasta el receptor de referencia atraviesa los mismos obstáculos atmosféricos que la señal que llega hasta el receptor móvil.

Contra las interferencias en la señal GPS

Las emisiones que radia el GPS son muy débiles. Para poder utilizarlas es esencial que el espectro de radio sea muy silente: incluso fuentes débiles de radiofrecuencia pueden estorbar o impedir el funcionamiento del GPS. La Comisión Federal de Comunicaciones de los EE.UU. ha establecido que las bandas destinadas al GPS han de mantenerse en total silencio. Solamente está presente el ruido electromagnético natural.

A pesar de estos esfuerzos, los usuarios para quienes la seguridad es de importancia fundamental —el caso del control de tráfico aéreo— se hallan expuestos a pérdidas de señal, sea por interrupciones accidentales o por emisiones malévolas destinadas a interferir el GPS. Felizmente, estos usuarios tienen acceso a un número cada vez mayor de defensas contra las interferencias. Los usuarios en vuelo, por ejemplo, pueden confiar en otros sistemas de navegación, basados en mediciones inerciales, en Loran-C o en equipos de telemetría. Las aplicaciones militares del GPS se valen a menudo de antenas inteligentes muy direccionales, capaces de desechar selectivamente las señales que interfieran (suprimiendo la recepción en ciertas direcciones) sin degradar apreciablemente la intensidad de la señal GPS. En un futuro no muy lejano, las aplicaciones para el mercado de consumo pueden muy bien aumentar la fiabilidad del GPS con mediciones telemétricas a las antenas de estaciones de televisión cercanas o a estaciones base de telefonía móvil.

Señales más nítidas e intensas

A partir del año 2005, los satélites GPS empezarán a radiodifundir nuevas señales, que aumentarán la robustez de los servicios y contribuirán al ajuste fino de su precisión localizadora al eliminar los errores ionosféricos (*véase el recuadro* “Múltiples señales, para mayor fiabilidad”). A las bandas L1 y L2 les serán añadidas dos señales militares, y a la L2 una civil. Las señales actuales seguirán activas, para garantizar que los receptores existentes funcionarán en el futuro. Allá por 2008, una serie adicional de satélites GPS perfeccionados comenzará a emitir todavía más señales para usos civiles en una tercera banda de frecuencias llamada L5. (Las bandas L3 y L4 portan información de carácter militar, sin relación con la navegación.) Las nuevas señales L5 serán cuatro veces más potentes que las actuales.

Estas señales complementarias permitirán que un único receptor pueda calcular el retraso de transmisión provocada por la ionosfera, con la consiguiente reducción de errores. Las señales L1, que viajan a través de la irregular capa ionosférica, experimentarán distinto retraso que, por ejemplo, las enviadas por L5. Los receptores futuros podrían, pues, comparar los retrasos en las señales recibidas por L1 y por L5. Podrían después utilizar este cálculo para estimar la densidad de electrones de la ionosfera y corregir sus efectos. Este es el cálculo que algunos costosos equipos GPS de uso civil intentan efectuar a partir de la señal civil de L1 y la señal militar de L2. Dado que las señales civiles utilizarán códigos de conocimiento público, desaparecerán las flaquezas que en la actualidad padece el procesado de señales de dos frecuencias. Ello significa que los receptores de funcionamiento en dos frecuencias, incluso en tres, serán habituales para usos civiles comerciales, hasta en equipos de uso personal.

También los operadores de unidades de GPS diferencial se beneficiarán del nuevo sistema de señales. Recordemos que la precisión del GPS-D se degrada con la distancia del usuario al receptor de referencia, porque el haz radiado desde el satélite hasta el usuario atraviesa la ionosfera en un punto cada vez más alejado del lugar por donde el haz de referencia traspasó la capa de plasma. Al disponer de varias frecuencias, el receptor móvil podrá evaluar la ionosfera de forma autónoma y las correcciones del GPS diferencial valdrían para mitigar los otros errores (de menor monto). Los futuros usuarios del GPS-D alcanzarán precisiones de 30 a 50 centímetros.

Los usuarios más exigentes del sistema actual, topógrafos, científicos o agricultores, necesitan precisiones centimétricas y, en ocasiones, milimétricas. Una precisión semejante requiere una forma avanzada

de GPS diferencial que vaya más allá de la aplicación de códigos de RPA, una técnica avanzada que mire debajo de esos códigos y mida el tiempo de llegada de las propias ondas portadoras que transportan las señales de GPS desde la órbita.

Las ondas de radiofrecuencia portadoras de las señales de GPS son microondas sinusoidales. Un ciclo individual tiene una longitud de onda —la distancia entre crestas— de 19 centímetros. Los receptores pueden medir el tiempo de llegada con una precisión que ronda el 1 por ciento. Tal resolución corresponde a una distancia de viaje de uno o dos milímetros, que es el grado de resolución que necesitan los usuarios más exigentes. Ahora bien, las mediciones efectuadas sobre la onda portadora son ambiguas; es decir, el receptor no puede decidir de qué ciclo se trata. A menos que el equipo pueda identificar de forma inequívoca el ciclo individual que está rastreando, la medición puede sufrir un error igual a cualquier número de longitudes de onda completas.

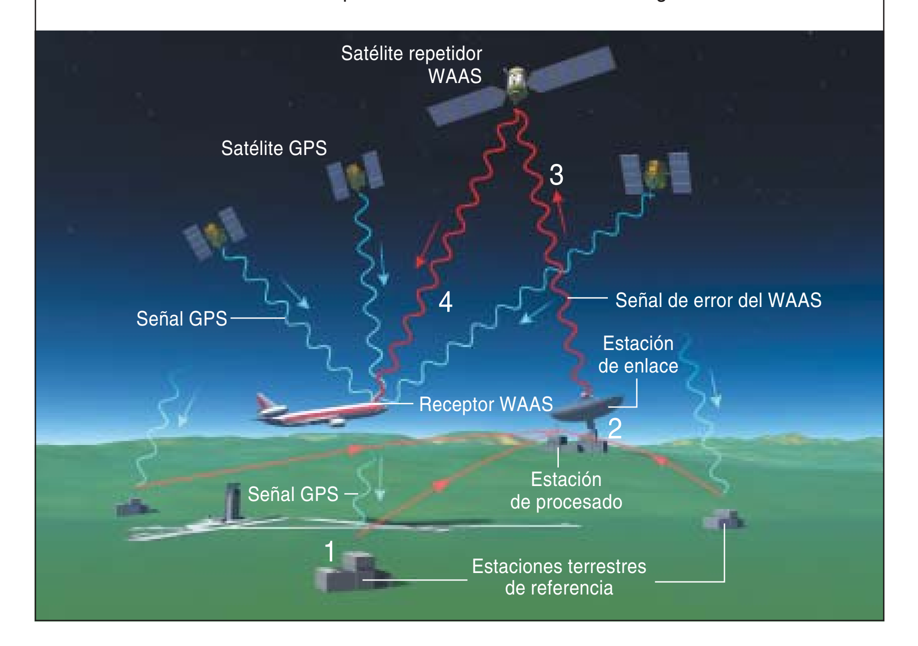
Esta dificultad recuerda la de medir una distancia utilizando las divisiones más finas de una regla. A diferencia de las divisiones gruesas, las finas están muy cercanas entre sí y permiten mediciones precisas, pero como no están rotuladas individualmente, son ambiguas. Por fortuna, existe un procedimiento que vincula sin ambigüedad la resolución burda de la escala de 30 centímetros del GPS diferencial con la fina resolución de la escala de dos milímetros de las longitudes de onda portadoras. Este proceso genera una escala de medición intermedia con una resolución adecuada para conectarlas. El puente computacional que cubre las mediciones se halla en esta escala intermedia, como explicaré seguidamente.

El problema resulta más fácil de comprender mediante un símil. Como ya se ha mencionado antes, los códigos RPA vienen a ser como un estribillo de una canción en el que cada nota fuese diferente e identificable. Las mediciones de la onda portadora se parecerían a la percusión que acompaña la canción, que marca muchos tiempos por nota. Si solamente oímos los golpes de la batería, nos resultará difícil saber qué parte de la canción estamos escuchando. La clave consiste en utilizar las notas de la canción para identificar de qué golpe se trata. En el caso del GPS resulta una tarea difícil: el instante de comienzo de cada nota (chip de código de PRN) puede determinarse con una precisión de sólo 30 centímetros. Cada golpe de la percusión (la longitud de onda de la portadora) dura tan sólo 19 centímetros. Con esa separación, están demasiado juntos para poder discernirlos; no es posible separarlos con la precisión que permiten los 30 centímetros de los códigos PRN.

Para identificar un golpe individual es necesario un tambor adicional, que vaya marcando un ritmo más lento. Los receptores de GPS avanzados crean este ritmo más lento multiplicando las portadoras L1 y L2 para producir una “frecuencia de batido”. También esta operación tiene una analogía musical. Al tocar dos notas simultáneamente en un instrumento, el oyente percibe los tonos originales, pero capta también uno nuevo, correspondiente a la diferencia de las frecuencias originales: la frecuencia de batido. Dado que la nueva frecuencia es igual a la diferencia de las anteriores, es necesariamente menos aguda que cualquiera de las dos notas originales. La longitud de onda es inversamente proporcional a la frecuencia; una frecuencia más baja entraña una longitud de onda mayor. En GPS, la longitud de onda asociada a la frecuencia diferencia es de 85 centímetros, y el sistema puede medirla con una resolución de unos 8 milímetros. Esta longitud de onda es lo bastante larga para que se la pueda

EL VUELO CON WAAS

La seguridad de los vuelos es mucho mayor cuando los pilotos conocen con exactitud la posición de sus aviones. El sistema de afinación del GPS para grandes áreas (WAAS), diseñado por la Administración Federal de Aeronáutica (FAA) de los EE.UU., mejora la exactitud y la integridad de las señales GPS de importancia vital. El WAAS ofrece precisiones de uno a dos metros en el eje horizontal y de dos a tres metros en el vertical sobre casi todo el territorio estadounidense. El sistema se basa en una red de estaciones de referencia instaladas en tierra, en ubicaciones conocidas (*1*). Cada estación compara los datos de los satélites del GPS con sus propias y confirmadas coordenadas geográficas y desarrolla correcciones para todos los satélites que están a la vista, que seguidamente son transferidas a uno de los puntos centrales de procesamiento (*2*). Los datos de corrección son enviados a satélites repetidores geoestacionarios (*3*), que, a su vez, los radiodifunden a los receptores WAAS (*4*), los cuales descodifican en “tiempo real” las correcciones de la geolocalización.



ALFRED T. KAMAJIAN

resolver con la precisión de 30 centímetros de las mediciones de código del receptor. Así pues, los receptores caros que emplean esta técnica satisfacen los requisitos de los usuarios más exigentes.

Este puente computacional entre los chips de código RPA y los ciclos de la portadora subyacentes quedará muy reforzado con las nuevas señales GPS que pronto entrarán en servicio. Los receptores de uso civil tendrán acceso a códigos públicos en la señal L2, así como en la nueva señal L5. Los receptores podrán, asimismo, procesar un trío de notas de batido (L1 menos L2, L1 menos L5, L2 menos L5), que proporcionarán varias sendas desde los chips de código RPA hasta los ciclos de portadora, y con ello geolocalizaciones de una precisión elevadísima.

El GPS en tiempo de guerra

A lo largo de estos últimos años, el número de usuarios civiles del GPS ha llegado a superar abrumadoramente al de usuarios militares, para quienes fue inicialmente construido el sistema. La señal civil está disponible para quienquiera posea un receptor GPS. Pero dado que las fuerzas armadas de EE.UU. dependen del GPS para la navegación y para fijar las coordenadas de los objetivos, el uso militar se vuelve prioritario cuando hay amenaza de conflicto bélico. En las regiones del mundo donde hay guerra abierta, los EE.UU. pueden "cegar" el GPS en la región emitiendo intensas señales justo en el centro de las bandas a que pertenecen las débiles señales GPS. A pesar de ello, el uso militar autorizado continúa en servicio, porque las emisiones de señales militares están lo suficientemente alejadas del centro de las bandas de uso civil y no resultan afectadas. En estas circunstancias, a sus adversarios les resulta imposible la utilización de las señales militares, porque los códigos militares son secretos. Por otra parte, las interferencias que el otro ejército pudiera llegar a crear en las señales militares del GPS serían, muy probablemente, efímeras, ya que las fuerzas armadas estadounidenses pueden detectar rápidamente y destruir las fuentes de la interferencia, como se demostró en la reciente guerra de Irak. El uso civil del GPS podría continuar fuera de la zona de conflicto, porque cualquier señal interferidora habría perdido casi toda su potencia a gran distancia de su emisor.

Un GPS para la navegación aérea

Para ver algunas de las consecuencias que el GPS perfeccionado podrá tener en el mundo real, fijémonos en las nuevas técnicas de ayuda a la navegación aérea de la Administración Federal de Aeronáutica (FAA) de Estados Unidos. Ahí, claro está, la fiabilidad reviste una importancia crítica. Los sistemas más innovadores, algunos de los cuales están ya parcialmente en servicio, permitirán a las tripulaciones servirse del GPS para guiar los aviones directamente hasta las pistas de aterrizaje en condiciones meteorológicas adversas, incluso con visibilidad nula. Para llevar a cabo esta tarea con certeza y sin riesgos, se requiere mucho más que mera precisión en la navegación. En primer lugar, las tripulaciones han de conocer cuál es el máximo error posible de su localización (es decir, una barra de error) en todas las circunstancias. Al efectuar la maniobra de aproximación final, por ejemplo, los pilotos no pueden admitir errores de más de 10 metros. En segundo lugar, los usuarios necesitan tener la seguridad de que sus sistemas de navegación no sufrirán interrupciones.

La FAA ha desarrollado dos sistemas basados en el GPS diferencial que proporcionan en tiempo real estas cotas máximas de error para los datos de localización. Cuentan con redes de receptores de referencia que supervisan de forma continua las mediciones GPS, aunque operan con independencia del segmento de control en tierra.

El sistema WAAS (*Wide-Area Augmentation System*, sistema de afinación del GPS para grandes áreas) entró en servicio en 2003 en los EE.UU. Se basa en una red de sensores repartidos por todo el país, destinados a medir el funcionamiento del GPS (*véase el recuadro* "El vuelo con WAAS"). Estos monitores se parecen a las estaciones de referencia del GPS diferencial: también WAAS proporciona correcciones para aumentar la precisión; pero, además, compara las correcciones de localización correspondientes a varias estaciones con el fin de generar las cotas de error, que son cruciales para el guiado de aeronaves. A continuación se vale de satélites geoestacionarios para trasladar a las tripulaciones esa información. En caso necesario, WAAS puede ajustar el margen de error transmitido en 7 segundos. El sistema detecta con precisión la ubicación de los aviones en vuelo y ayuda a conducir a los que estén en la maniobra de descenso hacia los aeropuertos hasta una altitud de 100 metros. Se está trabajando en sistemas parecidos en Europa (EGNOS), China, Japón, India, Australia y Brasil.

Donde el sistema WAAS abandona, toman el relevo los sistemas locales, que se encargan del control de los aviones en los segmentos inferiores de sus trayectorias de aterrizaje. Con el tiempo, el sistema de afinación del GPS en área local (LAAS) hará posibles aterrizajes completamente automáticos con visibilidad nula. Dado que sólo presta servicio a los aviones próximos a un aeropuerto, utiliza un enlace de radio de corto alcance para enviar las correcciones y las cotas de error. LAAS guarda estrecha relación con el sistema JPALS (*Joint Precision Approach and Landing System*, sistema conjunto de precisión para la aproximación y aterrizaje), aún en desarrollo, que servirá para guiar a los aviones hasta las cubiertas de los portaaviones, pese a sus cabeceos y balanceos. Los aviadores, en la aproximación final, han de controlar la altitud de sus aparatos con respecto a una cubierta móvil con un error de menos de un metro; si no, no podrían estar seguros de que el gancho de cola que pende del fuselaje quedará prendido en el cable de captura.

Los ingenieros de la marina están tratando de hacer que los aterrizajes en portaaviones sean más sencillos y más seguros mediante JPALS, que sitúa el receptor de referencia del GPS diferencial en el portaaviones. El sistema deberá entrar en fase de pruebas este mismo año. Tanto LAAS como JPALS son sistemas de doble frecuencia, ya que

se requieren dos frecuencias para garantizar la precisión durante esas operaciones que tanta necesitan. El sistema JPALS se servirá de las señales de uso militar disponibles en la actualidad en L1 y L2.

Si bien todos estos perfeccionamientos citados harán al GPS casi ubicuo, el gobierno estadounidense ha empezado a planificar la próxima ronda de perfeccionamientos de la técnica de navegación por satélite, o GPS III. Se buscan mayores fiabilidades, precisiones y robusteces frente a interferencias ocasionales y perturbaciones deliberadas, así como alentar la adopción de otros servicios de geolocalización y de aplicaciones nuevas y más elaboradas, como los sistemas inteligentes de seguridad viaria. Varios grupos empresariales, entre ellos Boeing y la asociación de Lockheed Martin y Spectrum Astro, han anunciado que competirán por la adjudicación de los contratos de miles de millones de dólares. El primer satélite GPS III podría lanzarse a comienzos del próximo decenio.

El autor

Per Enge es profesor de aeronáutica y astronáutica en la Universidad de Stanford y director del Laboratorio de Investigación del GPS, donde trabaja en la provisión de cotas de error en tiempo real de la localización por GPS y en la protección contra las interferencias de radiofrecuencia.

Bibliografía complementaria

PROPOSED NEW L5 CIVIL GPS CODES. J. J. Spiker, Jr. y A. J. Van Dierendonck en *Navigation Journal of the Institute of Navigation*, vol. 48, n.º 3, págs. 135-143; otoño de 2001.

GLOBAL POSITIONING SYSTEM: SIGNALS, MEASUREMENTS, AND PERFORMANCE. Pratap Misra y Per Enge. Ganga-Jamuna Press, 2001. Disponible en Navtech Seminars.

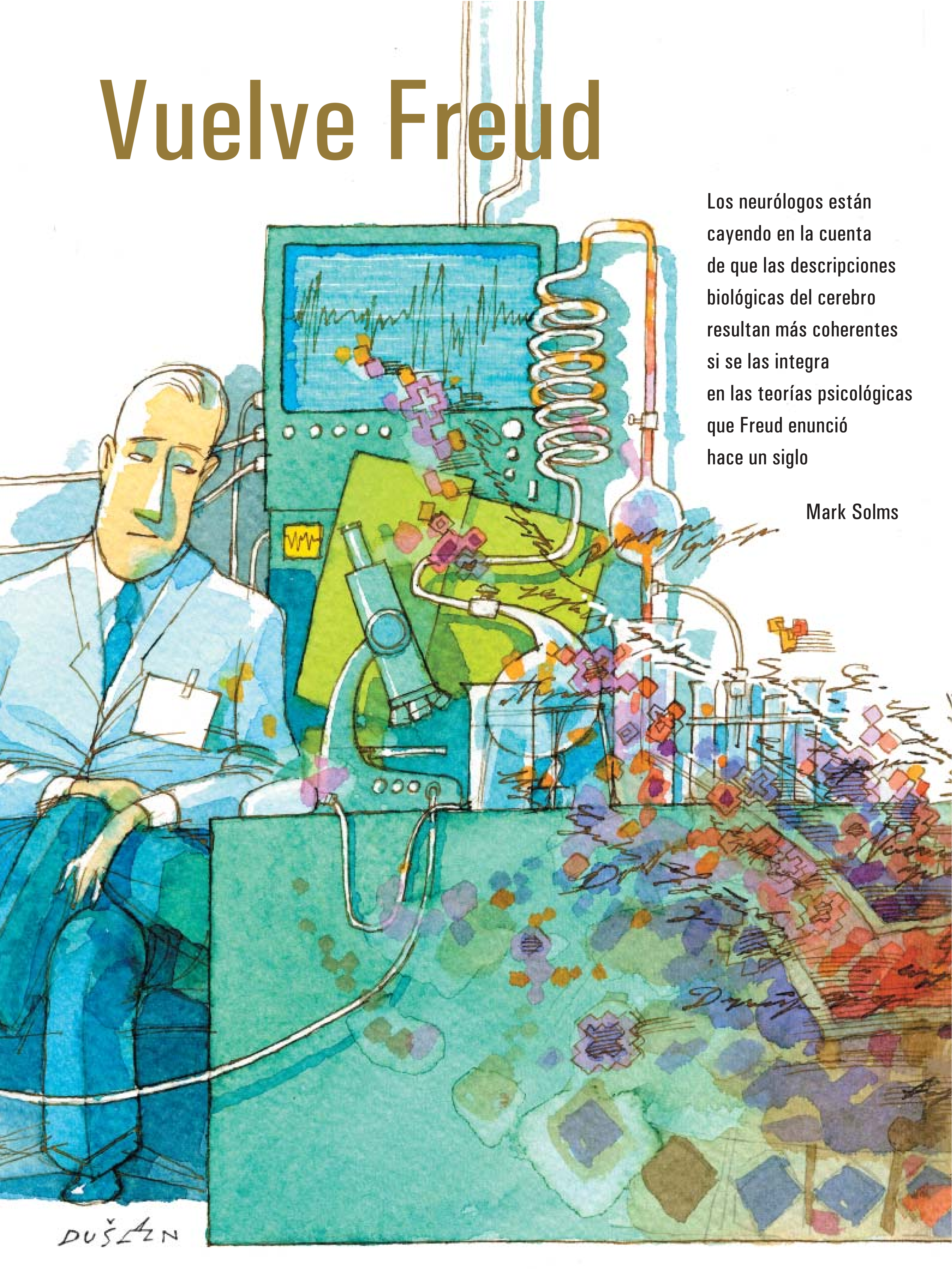
THE ROSUM TELEVISION POSITIONING TECHNOLOGY. M. Rabinowitz y J. J. Spiker en *Proceedings of the 59th Annual Meeting of the Institute of Navigation*; 2003.

GPS WORLD. Revista mensual publicada por Advanstar Communications desde 1999 hasta 2002.

Vuelve Freud

Los neurólogos están cayendo en la cuenta de que las descripciones biológicas del cerebro resultan más coherentes si se las integra en las teorías psicológicas que Freud enunció hace un siglo

Mark Solms



Durante la primera mitad del siglo XX, las ideas de Sigmund Freud predominaban cuando se trataba de explicar la naturaleza de la mente o psique humana. Su tesis básica decía que nuestras motivaciones permanecen en muy buena medida ocultas en nuestro inconsciente. Más aún, una fuerza represiva las aparta activamente de la consciencia. El aparato activo de la psique (el "yo") rechaza todos aquellos impulsos inconscientes (el "ello") que puedan promover una conducta incompatible con la civilizada idea que nos hacemos de nosotros mismos. Tal represión es necesaria porque los impulsos instintivos se expresan en forma de pasiones incontroladas, fantasías infantiles y deseos sexuales y agresivos.

Hasta su muerte en 1939, sostuvo que la enfermedad mental se produce cuando falla la represión. Las fobias, los ataques de pánico y las obsesiones se deberían a que en la conducta voluntaria se introducen furtivamente pulsiones ocultas. Correspondía, pues, a la psicoterapia rastrear los síntomas neuróticos hasta dar con sus raíces inconscientes y exponerlas al juicio maduro, racional, quitándoles así su fuerza compulsiva.

Sin embargo, al ir adquiriendo las investigaciones sobre la mente y el cerebro mayor complejidad a partir del decenio de 1950, a los especialistas les fueron pareciendo poco sólidas las pruebas que Freud adujo en pro de sus teorías. No las sometía a la experimentación controlada; sólo entreveraba inferencias teóricas y observaciones simples de pacientes en entornos clínicos.



1. FREUD JOVEN, hacia 1891.

Los tratamientos por medio de drogas ganaron terreno y la tendencia a explicar biológicamente las enfermedades mentales fue desprestigiando al psicoanálisis. Freud, de haber vivido por aquellos años, hasta puede que hubiese dado por bueno el cambio de enfoque. En su época, contó con una reputación como neurólogo muy alta. Hizo a menudo comentarios de este tipo: "Es de creer que las deficiencias de nuestra teoría se subsanarían si fuésemos capaces de sustituir los términos psicológicos por formulaciones fisiológicas y químicas". Pero no disponía aún de la ciencia o la técnica necesarias para conocer la organización y las funciones del cerebro, ni de la personalidad normal ni de la neurótica.

A la altura del decenio de 1980, hasta en algunos círculos psicoanalíticos se consideraba que las nociones del yo y el ello estaban irremediablemente anticuadas. Freud era ya historia. En la nueva psicología prevaleció la idea de que los deprimidos padecían su mal no porque algo socavase sus primeros afectos en la infancia, sino a causa de desarreglos de los componentes químicos del cerebro. La psicofarmacología, sin embargo, no proporcionó una teoría general de la personalidad, la emoción y la motivación, una concepción nueva de "lo que nos conmueve", que sustituyese a la de Freud. Sin un modelo así, los neurólogos restringieron su enfoque y prescindieron de las visiones de conjunto.

Hoy día se está volviendo a la visión de conjunto, y lo sorprendente es que el cuadro resultante se parece al que hace un siglo pergeñó Freud. Aún se está lejos del consenso, pero cada vez son más los neurólogos que llegan a la misma conclusión que Eric R. Kandel, de la Universidad de Columbia, premiado el año 2000 con el Nobel de fisiología o medicina: que "acerca de la mente, no hay todavía concepción más coherente e intelectualmente satisfactoria que el psicoanálisis".

Freud vuelve, pues, y no sólo a la teoría. En casi todas las grandes ciudades del mundo se han formado grupos de trabajo interdisciplinario que unen los campos, antes divididos y con frecuencia antagónicos, de la neurología y del psicoanálisis. Estas redes de investigación, a su vez, han constituido la Sociedad Internacional Neuropsicoanalítica, que organiza un congreso anual y publica la prestigiosa revista *Neuro-Psychoanalysis.* Testimonia el renovado respeto por las ideas de Freud el consejo editorial de esta revista, donde figura la plana mayor de la actual neurología del comportamiento: Antonio R. Damasio, Kandel, Joseph E. LeDoux, Benjamin Libet, Jaak Panksepp, Vilayanur S. Ramachandran, Daniel L. Schacter y Wolf Singer.

Estos investigadores están forjando, en palabras de Kandel, "un nuevo marco intelectual para la psiquiatría". Dentro de este marco, parece que la organización general de la mente que a grandes rasgos esbozó Freud está destinada a desempeñar un papel semejante al que representa la teoría de la evolución de Darwin para la genética molecular: una plantilla sobre la que se podrán ir colocando coherentemente los detalles que se vayan descubriendo. Los neurólogos están encontrando pruebas que avalan algunas de las teorías de Freud y, a la vez, van atando cabos acerca de los mecanismos subyacentes a los procesos mentales que describió.

Resumen/Teorías sobre el psiquismo

- Durante décadas, los conceptos freudianos —el yo, el ello, los deseos reprimidos— dominaron la psicología y los esfuerzos psiquiátricos por curar las enfermedades mentales. Pero una mejor comprensión de la química del cerebro fue sustituyendo este modelo por una explicación biológica de cómo se originan las funciones psíquicas en la actividad neuronal.
- Sin embargo, las más recientes tentativas de conjuntar diversos descubrimientos neurológicos van conformando una concepción química de la mente que revalida el esquema general propuesto por Freud hace casi un siglo. Hoy en día, cada vez es más numeroso el grupo de científicos dispuestos a conciliar la neurología y la psiquiatría dentro de una teoría unificada.

Motivación inconsciente

Cuando Freud introdujo su idea fundamental, que la mayoría de los procesos mentales que a diario determinan nuestros pensamientos, sentimientos y voliciones ocurren inconscientemente, sus contemporáneos la rechazaron por imposible. Pero hoy las investigaciones

DUSAN PETRICIC (*páginas anteriores*); BETTMANN/CORBIS

confirman cada vez más la existencia y la importante función de los procesos mentales inconscientes. Por ejemplo, la conducta de pacientes incapaces de recordar conscientemente sucesos acaecidos después de que se les dañaran ciertas estructuras cerebrales codificadoras de la memoria se halla a las claras influida por esos sucesos "olvidados". Los neurólogos cognitivos interpretan estos casos bosquejando dos sistemas mnémicos diferentes, uno que procesa la información "de modo explícito" (conscientemente) y otro que la trata "de modo implícito" (inconscientemente). Pues bien, Freud dividió la memoria de la misma manera.

Se han identificado también sistemas de memoria inconsciente que intervienen en el aprendizaje emocional. En 1996, en la Universidad de Nueva York, demostró LeDoux que bajo el córtex consciente hay una vía neuronal que conecta las informaciones de la percepción con las primitivas estructuras del cerebro que generan las reacciones de temor. Como esta vía elude el hipocampo —generador de los recuerdos conscientes—, los acaecimientos presentes suscitan de ordinario recuerdos inconscientes de sucesos pasados que tuvieron importancia emocional; por eso se producen sentimientos conscientes que parecen irracionales, como, digamos, una animadversión hacia quienes lleven barba.

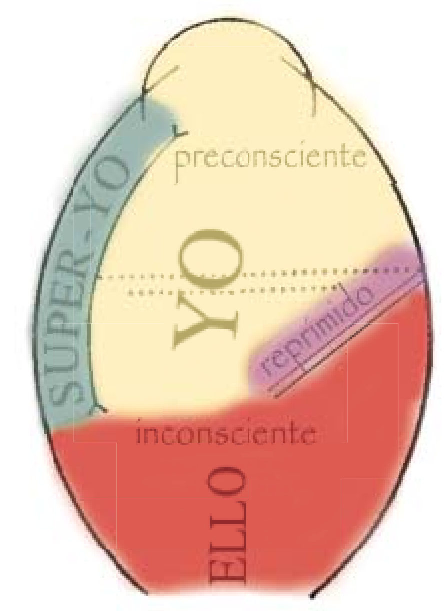
La neurología ha demostrado que las estructuras cerebrales principales de las que depende la formación de los recuerdos conscientes (explícitos) no funcionan durante los dos primeros años de la vida, proporcionando con ello una elegante explicación de lo que Freud llamaba amnesia infantil. Tal y como Freud conjeturó, no es que olvidemos nuestras primeras impresiones mnémicas, sino simplemente que no podemos evocarlas, hacerlas venir a la consciencia. Pero esta impotencia no impide que aquellas impresiones afecten a los sentimientos y a la conducta del adulto. Sería difícil encontrar un neurobiólogo experimental que discrepe de que las primeras experiencias, sobre todo las que se dan entre una madre y su bebé, influyen en las conexiones cerebrales hasta el punto de configurar los fundamentos de nuestra personalidad y nuestra futura salud mental. Y, sin embargo, ninguna de tales experiencias puede ser conscientemente rememorada. Se va haciendo cada vez más claro que gran parte de nuestra actividad mental está motivada inconscientemente.

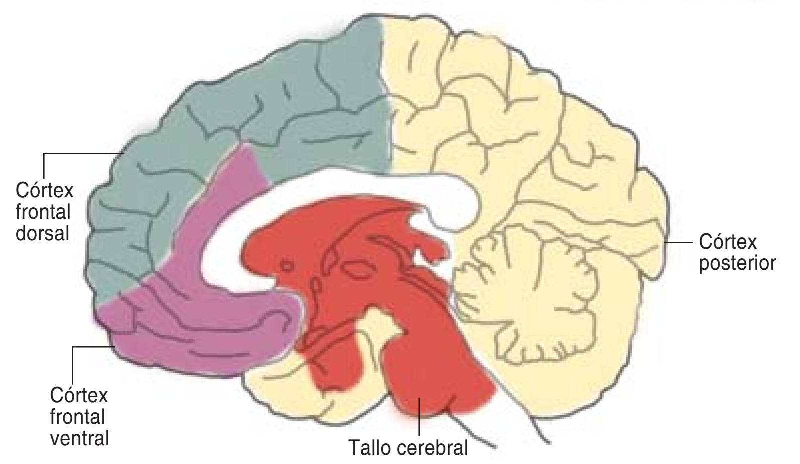
Se reivindica la represión

Sin embargo, aun suponiendo que la mayoría de las veces nos mueven pensamientos inconscientes, no quedaría probada la tesis de Freud según la cual reprimimos activamente informaciones insoportables. Pero cada día son más numerosos los estudios de casos que abonan tal idea. El más famoso fue el realizado en 1994 por el neurólogo conductual Ramachandran, de la Universidad de California en San Diego, con anosognósicos. A esos pacientes el deterioro de la región parietal derecha del cerebro les impedía tener consciencia de sus propias, y muy graves, deficiencias físicas, la parálisis de un miembro por ejemplo. Observó Ramachandran que una mujer adquirió de pronto conciencia, tras activarle artificial-

MENTE Y MATERIA

Freud elaboró su modelo definitivo de la mente en 1933 (*a la derecha*; *se han añadido colores a su esquema*). Las líneas de puntos representaban el umbral entre los procesos inconscientes y los conscientes. El 'super-yo' reprimía las pulsiones instintivas (el "ello") e impedía así que trastornaran el pensamiento racional. La mayoría de los procesos racionales ('el yo') eran automáticos e inconscientes, de modo que para dirigir la experiencia consciente, estrechamente vinculada a la percepción, sólo quedaba una pequeña parte del yo (*bulbo superior*). El 'super-yo' mediaba en la lucha que por la dominación mantenían el 'yo' y el 'ello'. La cartografía neurológica reciente (*abajo*) tiene bastante correlación, en líneas generales, con la concepción freudiana. Los centrales tallo cerebral y sistema límbico —responsable de los instintos y pulsiones— corresponden aproximadamente al 'ello' de Freud. La región ventral del córtex frontal, controladora de la inhibición selectiva, la región dorsal frontal, controladora del pensamiento consciente de sí, y el córtex posterior, que rige las representaciones del mundo externo, equivalen al 'yo' y al 'super-yo'.





A. W. FREUD ET AL., POR ACUERDO CON PATERSON MARSH LTD., LONDRES (*arriba*); OLIVER TURNBULL (*abajo y coloreado*)

mente el hemisferio derecho, de que tenía paralizado el brazo izquierdo; cayó también en la cuenta de que lo había tenido así los ocho días pasados desde que sufrió un accidente cerebrovascular. Demostraba que era capaz de reconocer su impedimento y que lo había tenido inconscientemente registrado durante los ocho días anteriores, pese a que durante todo ese tiempo negara conscientemente que sufriese problema alguno.

Es significativo que, una vez desaparecidos los efectos de la estimulación, la mujer no sólo volvió a creer que su brazo estaba en condiciones normales, sino que además olvidó la parte de la entrevista en la que había reconocido que tenía el brazo paralizado, aunque sí siguió recordando los demás detalles de la entrevista. Ramachandran concluía: "De estas observaciones se sigue una conclusión: que es cierto que los recuerdos se pueden reprimir selectivamente. ... Fue ver [a esa paciente] lo que me convenció de la realidad de los fenómenos de represión que constituyen la piedra angular de la teoría psicoanalítica clásica".



2. CON UN ESCANEADO DEL CEREBRO se pueden hoy ver las lesiones de disfunciones psíquicas que Freud sólo podía estudiar clínicamente. En esta imagen del cerebro, tomada por resonancia magnética, de un paciente que fabula inventando grandiosas historias de su vida se nota (*flecha*) una lesión en el giro cingulado, la parte del lóbulo frontal medio cuyo funcionamiento normal, suponía Freud, impedía a los deseos inconscientes alterar la imagen racional que de sí mismo tuviera el sujeto.

Como les ocurre a los sujetos con "cerebro dividido", cuyos hemisferios han quedado separados —famosos gracias a los estudios que en las décadas de 1960 y 1970 realizó el hoy ya fallecido premio Nobel Roger W. Sperry, del Instituto de Tecnología de California—, los pacientes anosognósicos urden razones que les libren de tener que enfrentarse a lo que no querrían reconocer y se inventan explicaciones verosímiles pero irreales de sus actos inconscientemente motivados. De este modo, dice Ramachandran, el hemisferio izquierdo recurre de manera manifiesta a "mecanismos de defensa" freudianos.

Análogos fenómenos se ha demostrado ahora que se dan también en sujetos con el cerebro intacto. Según hizo notar el neuropsicólogo Martin A. Conway, de la Universidad de Durham, Inglaterra, en un comentario publicado el año 2001 en *Nature*, si en condiciones de laboratorio se pueden generar claros efectos de represión en individuos comunes, en situaciones traumáticas de la vida real debe de haberlos mucho más vivos.

El principio del placer

Pero Freud fue aún más lejos. No sólo dijo que mucha de nuestra vida mental es inconsciente y reprimida, sino también que la parte reprimida de la mente inconsciente sigue un principio distinto del "principio de realidad" por el que se rige el yo consciente. Este tipo de pensamiento inconsciente sólo se atiene a sus deseos y se salta alegremente las leyes de la lógica y del curso del tiempo.

De estar Freud en lo cierto, las lesiones de las estructuras cerebrales inhibidoras (sede del yo "represor") darían suelta a modos de actuación mental ilusorios, irracionales. Pues bien, esto es lo que se ha observado en pacientes que sufren algún deterioro de la región límbica frontal, que controla aspectos fundamentales de la consciencia de sí mismo. Algunos sujetos muestran un peculiar síndrome, la psicosis de Korsakov: no se percatan de que son amnésicos y colman las lagunas de su memoria con fabulaciones.

Aikaterini Fotopoulou, neuropsicóloga de Durham, estudió hace poco en mi laboratorio a uno de estos pacientes. En ninguna de las sesiones de 50 minutos celebradas en mi despacho durante doce días seguidos fue capaz aquel hombre de recordar que me había visto antes alguna vez, ni tampoco que le habían operado para extirpar de sus lóbulos frontales el tumor que le produjo la amnesia. En cuanto a él se refería, no había deficiencia alguna. Al preguntarle por la cicatriz de su cabeza, fabulaba explicaciones de lo más inverosímiles: que le había operado un dentista o que habían tenido que implantarle un *bypass* coronario. Era cierto que había sufrido, años antes, esas intervenciones; a diferencia de la del cerebro, habían sido enteramente satisfactorias.

De manera parecida, al preguntarle quién era yo y qué estaba haciendo él en mi laboratorio, sus respuestas variaban: yo era un colega, o un amigo con el que se iba de copas, o un cliente que le consultaba sobre asuntos de su especialidad profesional, o un compañero de equipo en un deporte que él no había vuelto a practicar desde que abandonara, docenas de años antes, la universidad, o un mecánico que reparaba uno de sus muchos coches deportivos (que en realidad nunca había tenido). Su actitud era también coherente con estas falsas creencias: recorría con la mirada toda la habitación en busca de su cerveza, o miraba por la ventana para ver el coche.

Al observador casual le llama la atención la cualidad desiderativa de estas falsas ideaciones; Fotopoulou ha confirmado objetivamente esa impresión mediante el análisis cuantitativo de una serie de 155 fabulaciones sucesivas del paciente. Sus falsas creencias no se debían a estimulaciones fortuitas, sino que nacían del "principio del placer", al que Freud atribuía un papel central en el pensamiento inconsciente. El hombre no hacía más que reformar o refundir la realidad como él deseaba que fuese. Otros investigadores, entre ellos Martin Conway, de Durham, y Oliver Turnbull, de la Universidad de Gales, han informado de observaciones semejantes. Aunque son

CORTESIA DE OLIVER TURNBULL

neurólogos cognitivos y no psicoanalistas, interpretan sus hallazgos en términos freudianos; en esencia, nos dicen que el daño de la región límbica frontal, la que produce las fabulaciones, deteriora los mecanismos de control cognitivo de los que depende la captación normal de la realidad y libera de la inhibición las implícitas pulsiones desiderativas que influyen en la percepción, en la memoria y en el juicio.

3. FREUD DIBUJO en 1895 este esquema de un mecanismo neuronal de la represión (*arriba*); esperaba que las explicaciones psicológicas del psiquismo serían un día remplazadas por explicaciones biológicas como ésa. En su esquema, un recuerdo desagradable sería activado normalmente por un estímulo ('Q_n' *en el extremo izquierdo)* que va de la neurona 'a' hacia la neurona 'b' (*abajo*). Pero la neurona 'alfa' (*a la derecha de* 'a') podía desviar la señal e impedir con ello la activación si otras neuronas (*arriba a la derecha*) ejercían un influjo "represor". Nótese que Freud (*aquí, en un momento muy posterior de su vida*) situó entre las neuronas unos hiatos que, predecía, servirían como "barreras de contacto". Dos años después, el fisiólogo inglés Charles Sherrington los descubrió y los llamó sinapsis.

El animal interior

Freud sostenía que el principio del placer da expresión a las pulsiones animales primitivas. Para sus victorianos contemporáneos, la implicación de que la conducta humana estuviese en el fondo regida por impulsos cuya más alta finalidad no era sino la plena satisfacción de los propios deseos carnales resultaba totalmente escandalosa. La indignación moral se fue debilitando durante los decenios subsiguientes, pero los investigadores del conocimiento fueron dejando de lado la visión freudiana del hombre como animal.

Ahora vuelve a estar en boga. Neurólogos como Donald W. Paff, de la Universidad Rockefeller, y Jaak Panksepp, de la Universidad estatal de Bowling Green, creen que los mecanismos instintivos que rigen la motivación humana son aún más primitivos de lo que Freud se imaginó. Los hombres compartimos con nuestros parientes los primates y con todos los mamíferos sistemas básicos de control de las emociones. En el profundo nivel de la organización mental que Freud llamaba el ello, la anatomía funcional y la química de nuestros cerebros no se diferencian mucho de las de nuestros animales domésticos.

Pero los neurólogos modernos no aceptan ya la manera en que Freud clasificó la vida instintiva humana, mediante una simple dicotomía entre la sexualidad y la agresividad. Gracias al estudio de los efectos de lesiones, drogas y estimulaciones artificiales del cerebro, han identificado en los mamíferos al menos cuatro circuitos instintivos básicos, algunos de los cuales se superponen: el sistema "anhelante" o "de recompensa" (que motiva la busca del placer); el sistema de "enfado-rabia" (que rige la agresión furiosa pero no la depredadora); el sistema de "temor-ansiedad"; y el sistema "pánico" (que incluye instintos complejos, así los que rigen los vínculos sociales). También se está investigando si existen más fuerzas instintivas, por ejemplo un díscolo sistema "lúdico". Neurotransmisores específicos (sustancias químicas que en el interior del cerebro llevan mensajes de unas neuronas a otras) modulan estos sistemas cerebrales.

El sistema anhelante, regulado por el neurotransmisor dopamina, guarda una notable semejanza con la "libido" freudiana. Según Freud, la pulsión libidinal o sexual es un sistema de busca de placer que activa enérgicamente la mayoría de nuestras interacciones con el mundo dirigidas a alcanzar objetivos. La investigación moderna patentiza que su equivalente neural interviene mucho en casi todas las formas de ansia y de adicción. Los experimentos que Freud efectuó en la primera parte de su carrera con la cocaína —sobre todo en sí mismo— le convencieron de que la libido ha de tener una base neuroquímica específica. A diferencia de sus sucesores, no vio que hubiese razón alguna para un antagonismo entre el psicoanálisis y la psicofarmacología. Predijo con entusiasmo que llegaría un día en el que "las energías del ello" serían controladas por "determinadas sustancias químicas". En la actualidad, los tratamientos que unen la psicoterapia y las medicaciones psicoactivas son ampliamente reconocidos como la mejor manera de hacer frente a muchas dolencias psíquicas. Y las técnicas de formación de imágenes cerebrales muestran que la terapia verbal afecta al cerebro de modo no muy diferente a como le afectan esos fármacos.

Los sueños tienen significado

Las ideas de Freud están volviendo a tener vigencia también en los estudios del dormir y los sueños. Su teoría de que las visiones nocturnas son vislumbres de deseos inconscientes quedó desacreditada cuando, en los años cincuenta, se descubrieron la fase 'REM' del sueño —o 'del movimiento rápido de los ojos'— y su estrecha correlación con los sueños. Y pareció perder toda credibilidad cuando, por los años setenta, se demostró que el ciclo del sueño estaba regulado por una sustancia química ubicua, la acetilcolina, que se producía en una parte del tallo

cerebral "ajena a la mente". El sueño REM ocurría automáticamente cada 90 minutos, poco más o menos, promovido por sustancias químicas y estructuras cerebrales que nada tenían que ver con la emoción ni con la motivación. Este descubrimiento implicaba que los sueños no significaban nada, sino que eran simplemente historias inventadas por el cerebro superior para tratar de reflejar la aleatoria actividad del córtex causada por el REM.

Pero investigaciones más recientes han revelado que el sueño onírico y el sueño REM son dos estados diversos, controlados por mecanismos distintos, aunque interactivos. Genera los sueños una red de estructuras cuyo centro son los circuitos instintivo-motivacionales del prosencéfalo o parte frontal del cerebro. Este descubrimiento ha originado numerosas teorías sobre el cerebro durmiente, en muchas de las cuales hay claras reminiscencias de la de Freud. Destaca sobre todo la observación, que otros y yo hemos hecho, de que cuando se cortan ciertas fibras del interior del lóbulo frontal, los sueños cesan por completo, síntoma que coincide con una reducción general del comportamiento motivado. Dicha lesión es justo la que se producía deliberadamente con la lobectomía frontal, obsoleto procedimiento quirúrgico al que se recurría antes para controlar las alucinaciones y los delirios. La sustituyó en los años sesenta el suministro de fármacos que aminoran la actividad de la dopamina en esos mismos sistemas cerebrales. El sistema anhelante podría ser, pues, el generador primario de los sueños. Esta posibilidad se ha convertido en un importante foco de la investigación actual.

De confirmarse la hipótesis, la teoría de que soñamos para satisfacer deseos reprimidos o suprimidos podría ocupar de nuevo un lugar central en las investigaciones sobre el sueño. Pero aun suponiendo que prevalezcan otras interpretaciones de los datos neurológicos más recientes, en conjunto vienen todas ellas a demostrar que las conceptualizaciones "psicológicas" del sueño tornan a ser científicamente respetables. Pocos neurocientíficos pretenden aún —como antaño, y sin réplica entonces— que el contenido de los sueños no responde a mecanismos emocionales primarios.

Acabar la tarea

Desde luego que no a todo el mundo le entusiasma la reaparición de las nociones freudianas, no ya en los márgenes, sino en la corriente principal de la investigación de la mente. A los psicoanalistas de más edad no les es fácil aceptar que sus colegas y discípulos más jóvenes puedan y deban hoy someter los saberes clásicos a un tipo de escrutinio biológico totalmente nuevo. Pero un alentador número de veteranos de ambas orillas del Atlántico se empeñan, por lo menos, en mantener una mentalidad abierta, como lo atestiguan los eminentes psicoanalistas del consejo editorial de la revista *Neuro-Psychoanalysis* y otros muchos que, entrados ya en años, forman parte de la Sociedad Internacional de Neuro-Psicoanálisis.

Lo que a los neurocientíficos les hace resistirse al retorno de las ideas psicoanalíticas es el fantasma de la mole freudiana que conocieron en los primeros años de sus carreras y que tan indestructible parecía. No pueden reconocer ni siquiera una confirmación parcial de las intuiciones fundamentales de Freud; piden una purga total (*véase el recuadro* "¿Vuelve Freud? Como un mal sueño"). Para J. Allan Hobson, famoso investigador del sueño y psiquiatra de la facultad de medicina de Harvard, el renovado interés por Freud es poco menos que un inútil vertido de datos modernos en un viejo odre conceptual. Pero, como decía Fred Guterl en una entrevista publicada en 2002 en *Newsweek*, para los neurólogos a los que entusiasma la reconciliación de la neurología con la psiquiatría "no se trata de demostrar si Freud estaba o no en lo cierto, sino de concluir la tarea".

Si puede ser llevada a cabo, si se logra establecer ese "nuevo marco intelectual para la psiquiatría" de que habla Kandel, quedará atrás el tiempo en que las personas con dificultades emocionales debían elegir entre la terapia verbal del psicoanálisis, ajena tal vez a la moderna medicina basada en pruebas empíricas, y la terapia con drogas prescrita por una psicofarmacología a la que quizá le falte tener en cuenta la relación entre la química cerebral sobre la que actúa y las complejas trayectorias de la vida real que culminan en el trastorno emotivo. La psiquiatría del mañana promete proporcionar a los pacientes una ayuda que se base en la comprensión integral del funcionamiento de la mente humana.

Sean cuales fueren las insospechadas terapias que aporte el futuro, el mejor conocimiento de lo que ocurre en el cerebro sólo puede beneficiar a los pacientes. Puesto que los neurocientíficos modernos abordan una vez más las profundas cuestiones de la psicología humana que tanto preocuparon a Freud, es gratificante caer en la cuenta de que podemos construir sobre los cimientos que él puso, en vez de tener que empezar todo de nuevo. Aun advirtiendo puntos débiles en las ricas teorías de Freud y tratando por ende de corregir, revisar y complementar su obra, nos sentiríamos privilegiados si lográsemos, en efecto, concluir la tarea.

El autor

Mark Solms es profesor titular de neuropsicología de la Universidad de Ciudad del Cabo (Sudáfrica) y profesor no titular honorario de neurocirugía en la Escuela Real de Medicina y Odontología de Londres. Es también director del Centro Arnold Pfeffer de Neuropsicoanálisis del Instituto Psicoanalítico neoyorquino y consultor neuropsicológico del Centro Anna Freud de Londres.
Agradece a Oliver Turnbull, profesor no titular del Centro de Neurología Cognitiva de la Universidad de Gales en Bangor, la ayuda prestada para la realización de este artículo.

Bibliografía complementaria

DREAMING AND REM SLEEP ARE CONTROLLED BY DIFFERENT BRAIN MECHANISMS. Mark Solms en *Behavioral and Brain Sciences*, vol. 23, n.º 6, págs. 843-850; diciembre, 2000.

FREUDIAN DREAM THEORY TODAY. Mark Solms en *Psychologist*, vol. 13, n.º 12, págs. 618-619; diciembre, 2000.

CLINICAL STUDIES IN NEURO-PSYCHOANALYSIS. K. Kaplan-Solms y M. Solms. Karnac Books, 2000.

THE BRAIN AND THE INNER WORLD. Mark Solms y Oliver Turnbull. Other Press, 2002.

¿VUELVE FREUD? COMO UN MAL SUEÑO

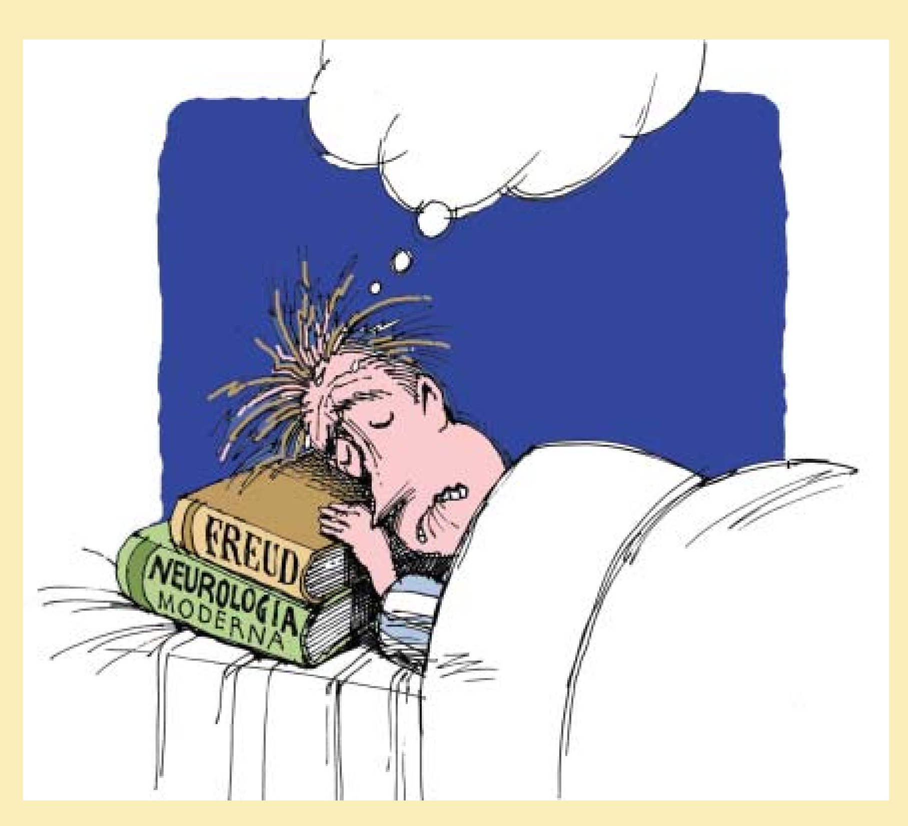
por J. Allan Hobson

Las opiniones de Sigmund Freud acerca del significado de los sueños constituían el núcleo de su teoría sobre el funcionamiento de la psique. Mark Solms y otros investigadores aseguran que los actuales estudios de imágenes y lesiones del cerebro están revalidando la concepción freudiana del psiquismo. Pero investigaciones científicas similares a ésas hacen ver que algunas de las principales tesis de Freud son seguramente erróneas.

Según Freud, la extraña naturaleza de los sueños se debería al complicado esfuerzo con que la psique trata de conciliar, mediante el disfraz simbólico y la censura, los inaceptables deseos instintivos que surgen del inconsciente cuando, al dormir, relaja el yo su prohibición del ello. Pero la mayor parte de los elementos de juicio empíricos de índole neurobiológica respalda más bien la opinión contraria: lo extraño de los sueños es debido a cambios normales del estado del cerebro, generados por unos mecanismos químicos del tallo cerebral que modulan la activación de varias regiones del córtex. Numerosos estudios han indicado que los cambios químicos determinan la cantidad y la cualidad de las visiones oníricas, de las emociones y de los pensamientos. La noción freudiana de 'disfraz-y-censura' se ha de desechar; nadie cree que la lucha entre el yo y el ello, si es que existe, controle la química del cerebro. La mayoría de los psicoanalistas no sostienen ya que el modelo disfraz-censura tenga validez.

Sin disfraz ni censura, ¿qué queda de la freudiana teoría del sueño? No gran cosa; tan sólo que las pulsiones instintivas podrían impeler la formación de los sueños. Ciertamente, las pruebas empíricas indican que activar las partes del sistema límbico que producen ansiedad, enfado y gozo configura los sueños; pero esas influencias no son "deseos". El análisis de los sueños muestra que en ellos las emociones son negativas tan a menudo como positivas, lo que querría decir que la mitad de nuestros "deseos" para con nosotros mismos son negativos. Y como cualquiera que haya soñado sabe, en los sueños las emociones raramente llevan disfraz: se introducen sin tapujos en la trama del sueño, produciendo con frecuencia efectos desagradables, pesadillas. Freud nunca fue capaz de explicar por qué son negativas tantas de las emociones oníricas.

Otro pilar de la teoría de Freud es que, como el verdadero significado de los sueños yace oculto, las emociones que reflejan sólo pueden revelarse por medio del tortuoso método de la libre asociación, conforme al cual el sujeto analizado ha de ir refiriendo cuanto le venga a las mientes, a la espera de dar con alguna conexión significativa. Mas este esfuerzo es innecesario, porque en realidad no hay tal ocultación del significado. En los sueños, lo que se ve es lo que hay. Las emociones quedan de manifiesto en el contenido de los sueños, y para advertir qué sentimientos representan éstos basta con que quienes los sueñan y sus terapeutas les presten la debida atención.



DUSAN PETRICIC

Solms y otros freudianos dan a entender que atribuir los sueños a la química equivale a negar que contengan mensajes emocionales. La teoría de la activación-síntesis química, que propusimos en 1977 Robert W. McCarley, de la facultad de medicina de Harvard, y yo, sólo mantenía que la explicación psicoanalítica de las peculiaridades oníricas por medio de un significado oculto era errónea. Nosotros hemos sostenido siempre que los sueños tienen carga emocional y significado.

Y ¿qué decir del sueño REM? Nuevos estudios revelan que se puede soñar fuera de la fase REM, pero en la teoría de la activación química nada impide que así ocurra; sólo establece que la frecuencia de los sueños es exponencialmente mayor en esa fase.

El psicoanálisis se halla en graves apuros, y ninguna componenda neurobiológica lo sacará de ahí. Tan radical es la revisión que necesita, que muchos neurocientíficos prefieren empezar de nuevo y crear un modelo neurocognitivo de la psique. La teoría psicoanalítica abarca mucho, sin duda, pero si sus errores son gravísimos, no cabe decir que esa amplitud sea una virtud. Los científicos que comparten esta opinión se declaran a favor de que se investiguen los sueños, las enfermedades mentales y la experiencia consciente normal por medio de modelos con mayor fundamento biológico que los ofrecidos por el psicoanálisis.

J. Allan Hobson, profesor de psiquiatría de la facultad de medicina de Harvard, ha escrito mucho sobre las bases cerebrales del psiquismo y sus implicaciones psiquiátricas. Para más detalles, véase su libro Dreaming: An Introduction to the Science of Sleep *(Oxford University Press, 2003).*

El universo antes de la gran explosión

Gabriele Veneziano

La teoría de cuerdas da a entender que la gran explosión no fue el origen del universo, sino sólo la evolución de un estado preexistente

¿Fue realmente la gran explosión el inicio del tiempo? ¿O el universo existía desde antes? Semejante pregunta parecía casi una blasfemia hace sólo una década. La mayoría de los cosmólogos sostenía que carecía de sentido, que contemplar un tiempo anterior a la gran explosión era como preguntarse por los lugares al norte del polo norte. Pero los desarrollos de la física teórica, especialmente la aparición de la teoría de cuerdas, han cambiado la perspectiva. El universo de antes de la explosión se ha convertido en la última frontera de la cosmología.

El nuevo afán de considerar lo que podría haber sucedido antes de la explosión es la última oscilación de un péndulo intelectual que ha ido y venido durante milenios. En una u otra forma, el problema del inicio ha atraído a filósofos y teólogos de casi todas las culturas. Está relacionado con un gran conjunto de interrogantes, uno de ellos famosamente resumido en un cuadro pintado en 1897 por Paul Gauguin: *D'où venons-nous? Que sommes-nous? Où allons-nous?* ¿De dónde venimos? ¿Qué somos? ¿Adónde vamos? La obra representa el ciclo del nacimiento, vida y muerte —origen, identidad y destino de cada individuo—; estos interrogantes personales conectan directamente con los cósmicos. Podemos remontarnos por nuestro linaje más allá de las generaciones, más allá de nuestros ancestros animales, más allá de las primitivas formas de vida y de protovida, más allá de los elementos sintetizados en el universo primordial, más allá de la energía amorfa que aún antes ocupaba el espacio. ¿Se extiende nuestro árbol genealógico indefinidamente hacia atrás? ¿O bien sus raíces acaban? ¿Es el cosmos pasajero, como nosotros?

Los antiguos griegos debatieron con ardor sobre el origen del tiempo. Aristóteles, de parte de quienes no pensaban que hubiera habido un inicio, invocó el principio de que de la nada, nada puede salir. Si el universo no pudo pasar de la nada a algo, es que había existido siempre. Por esta y otras razones, el tiempo debía extenderse eternamente en el pasado y en el futuro. Los teólogos cristianos mantenían el punto de vista opuesto. San Agustín afirmaba que Dios existe fuera del espacio y el tiempo, y puede llevarlos a la existencia de la misma manera que podía forjar otros aspectos de nuestro mundo. Cuando se le preguntaba qué hacía Dios *antes* de crear el mundo, Agustín respondía que, como el propio tiempo pertenece a la creación de Dios, no había un *antes*.

La teoría de la relatividad general de Einstein condujo a los cosmólogos modernos casi a la misma conclusión. La teoría mantiene que espacio y tiempo son "blandos", maleables. A las mayores escalas, el espacio es dinámico por naturaleza; se expande o se contrae con el tiempo, arrastrando la materia como la marea arrastra un tronco. Los astrónomos hallaron en el decenio de 1920 que nuestro universo se está, ahora, expandiendo: las galaxias distantes se separan unas de otras. Una consecuencia, como Steven Hawking y Roger Penrose probaron en los años sesenta, es que el tiempo no se puede extender indefinidamente hacia atrás. Si se proyecta la historia cósmica hacia atrás en el tiempo, todas las galaxias se juntan en un único punto infinitesimal, una "singularidad", casi como si estuvieran descendiendo en un agujero negro. Cada galaxia, o sus precursores, se comprime hasta un tamaño nulo. La densidad, la temperatura y la curvatura del espaciotiempo se hacen infinitas. La singularidad es el cataclismo primero, más allá del cual nuestra genealogía cósmica no se puede extender.

Extraña coincidencia

La singularidad inevitable pone en serios problemas a los cosmólogos. En particular, no encaja bien con el elevado grado de homogeneidad e isotropía que exhibe el universo a escalas grandes. Para que el cosmos parezca aproximadamente el mismo en todas partes, alguna clase de comunicación debe haber entre regiones distantes del espacio, para que se coordinen sus propiedades. Pero tal comunicación contradice el viejo paradigma cosmológico.

ALFRED T. KAMAJIAN

Para concretar, consideremos lo que ha sucedido durante los 13.700 millones de años transcurridos desde que se liberó la radiación del fondo cósmico de microondas. La distancia entre galaxias ha crecido en un factor de aproximadamente 1000 (a causa de la expansión), mientras que el radio del universo observable ha crecido en un factor mucho mayor, de alrededor de 100.000 (porque la luz es más veloz que la expansión). Hoy día vemos partes del universo que no podríamos haber visto hace 13.700 millones de años. En realidad, éste es el primer momento de la historia cósmica en que la luz de las galaxias más distantes ha llegado a la Vía Láctea.

Sin embargo, las propiedades de la Vía Láctea son básicamente las mismas que las de las galaxias distantes. Es como si se fuera a una fiesta y se comprobara que se lleva exactamente el mismo vestido que una docena de amigos. Si sólo dos estuvieran vestidos igual, se podría explicar como una coincidencia, pero una docena sugiere que los asistentes se han puesto de acuerdo. En cosmología, el número no es una docena, sino decenas de miles: el número de fragmentos de cielo independientes, aunque estadísticamente idénticos, del fondo de microondas.

Una posibilidad es que todas estas regiones del espacio estuvieran dotadas desde el inicio con idénticas propiedades; en otras palabras, que la homogeneidad sea mera coincidencia. Sin embargo, se han concebido un par de circunstancias más naturales que nos librarían de ese callejón sin salida: o el universo primitivo era mucho menor de lo supuesto en la cosmología ordinaria, o mucho más viejo. Una u otra (o ambas, actuando juntas) habrían posibilitado la intercomunicación.

La preferida sigue siendo la primera. El universo habría pasado por un período de expansión acelerada, o de "inflación", al comienzo de su historia. Antes de esta fase, las galaxias, o sus precursores, estaban tan juntas, que fácilmente podían coordinar sus propiedades. Durante la inflación, perdieron el contacto debido a que la luz no podía seguir el paso de la frenética expansión. Cuando la inflación acabó, la expansión empezó a desacelerarse, de manera que las galaxias, gradualmente, volvieron a verse unas a otras.

Se atribuye el empuje inflacionario a la energía potencial almacenada en un nuevo campo, el inflatón, unos 10^{-35} segundos después de la gran explosión. La energía potencial, al contrario que la masa en reposo o la energía cinética, conduce a la repulsión gravitatoria. En vez de ralentizar la expansión, como haría la gravitación de la materia ordinaria, el inflatón la acelera. Propuesta en 1981, la inflación ha explicado con precisión una amplia variedad de observaciones [*véase* "El universo inflacionario", por Alan H. Guth y Paul J. Steinhardt; INVESTIGACIÓN Y CIENCIA, julio de 1984]. Sin embargo, todavía quedan algunos problemas teóricos, empezando por el de qué es exactamente el inflatón y qué le dio una energía potencial inicial tan grande.

La segunda y menos conocida solución del problema prescinde de la singularidad. Si el tiempo no empezó en la explosión, si una larga era precedió al inicio de la presente expansión cósmica, la materia podría haber tenido todo el tiempo necesario para distribuirse homogéneamente por sí misma. De ahí que se haya reexaminado el razonamiento que llevó a inferir la existencia de una singularidad.

Una de sus hipótesis, que la teoría de la relatividad es válida siempre, resulta cuestionable. Cerca de la presunta singularidad, los efectos cuánticos deberían haber sido grandes, incluso dominantes. La relatividad ordinaria no tiene en cuenta tales efectos, de manera que aceptar la inevitabilidad de la singularidad equivale a fiarse de la teoría más allá de lo razonable. Para saber lo que realmente sucedió, hay que subsumir la relatividad en una teoría cuántica de la gravedad. La tarea ha venido ocupando a los teóricos desde la época de Einstein, pero el progreso fue casi nulo hasta mediados los años ochenta.

Evolución de una revolución

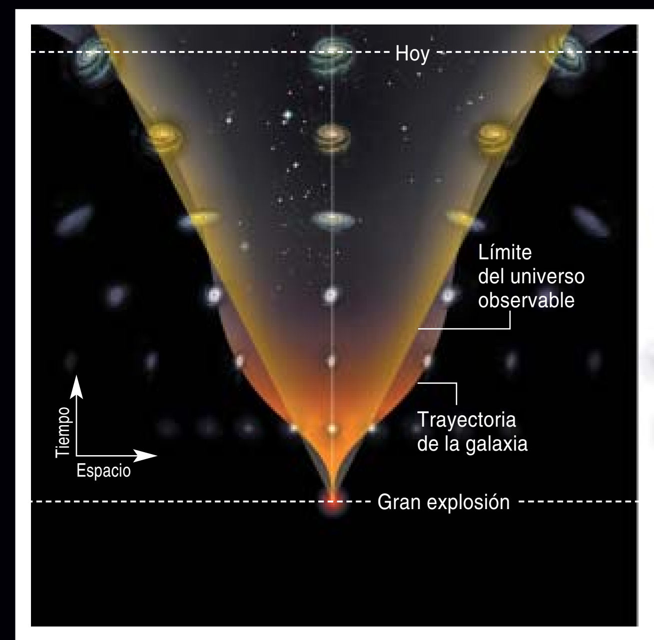
Hoy en día quedan dos enfoques: uno, la gravedad cuántica de bucles, mantiene la teoría de Einstein esencialmente intacta; pero cambia el procedimiento de llevarla a cabo en la mecánica cuántica [*véase* "Atomos del espacio y del tiempo", por Lee Smolin; INVESTIGACIÓN Y CIENCIA, marzo 2004]. Los seguidores de la gravedad cuántica de bucles han dado grandes pasos en los últimos años. A pesar de ello, su método podría no ser tan revolucionario como para resolver los problemas fundamentales de la cuantización de la gravedad. Un problema parecido afrontaron los teóricos de partículas después de que Enrico Fermi introdujera su teoría efectiva de la fuerza nuclear débil en 1934. Todos los esfuerzos por dar sentido a una versión cuántica de la teoría de Fermi fallaron. No se ne-

Resumen/Cosmología de cuerdas

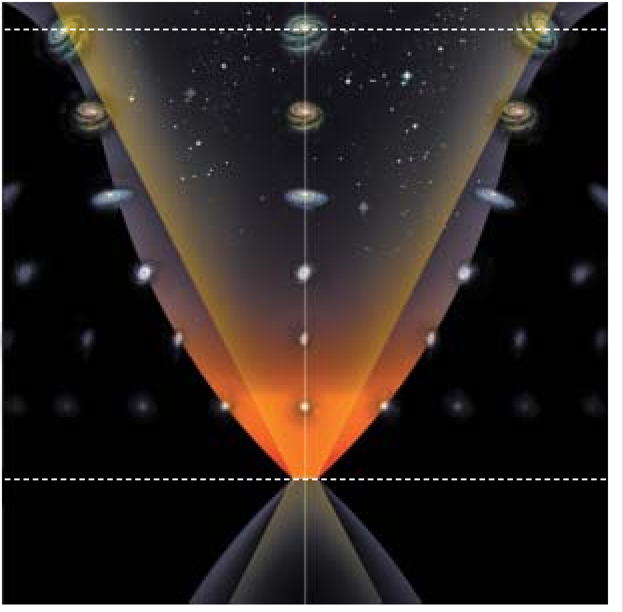
- Filósofos, teólogos y científicos siempre han debatido si el tiempo es eterno o finito; es decir, si el universo ha existido siempre o ha tenido en cambio una génesis definida. La teoría de la relatividad general de Einstein implica finitud. Un universo en expansión tiene que haber empezado en la gran explosión.
- Sin embargo, la relatividad general pierde su validez en la vecindad de la explosión, momento en que interviene la mecánica cuántica. El principal candidato actual a una teoría cuántica de la gravedad completa, la teoría de cuerdas, introduce un cuanto de longitud mínimo como una nueva constante fundamental de la naturaleza y convierte con ello en insostenible que la gran explosión fuera una verdadera génesis.
- La explosión tuvo lugar, pero no correspondió a un momento de infinita densidad. El universo podría ser anterior a ella. Las simetrías de la teoría de cuerdas sugieren que el tiempo no tuvo un inicio y no tendrá un final. Podría haber empezado casi vacío y haberse ido concentrando hasta la explosión, o incluso podría haber pasado por un ciclo de muerte y renacimiento. En cualquier caso, la época anterior a la explosión habría formado el cosmos actual.

Dos concepciones del inicio

EN NUESTRO UNIVERSO EN EXPANSION, las galaxias se alejan unas de otras. Cualquier par de galaxias se separa a una velocidad proporcional a la distancia que media entre ambas: si distan 500 millones de años-luz entre sí, se separarán el doble de deprisa que si distan 250. Por tanto, las galaxias deben haber partido del mismo lugar al mismo tiempo: la gran explosión. La conclusión vale incluso cuando la expansión cósmica ha pasado por períodos de aceleración y desaceleración; en los diagramas del espaciotiempo (*abajo*) las galaxias siguen trayectorias serpenteantes que entran y salen de la región observable del espacio (*cono amarillo*). Sin embargo, la situación deviene incierta en el preciso momento en que las galaxias (o sus antecesores) empezaron a separarse.



En la cosmología estándar de la gran explosión, que se basa en la teoría general de la relatividad de Einstein, la distancia entre dos galaxias cualesquiera era nula hace un tiempo finito. Antes de dicho momento, el tiempo pierde sentido.



En modelos más complejos que incluyen efectos cuánticos, dos galaxias cualesquiera deben haber partido a una cierta distancia mínima una de otra. Estos modelos abren la posibilidad de un universo anterior a la explosión.

SAMUEL VELASCO

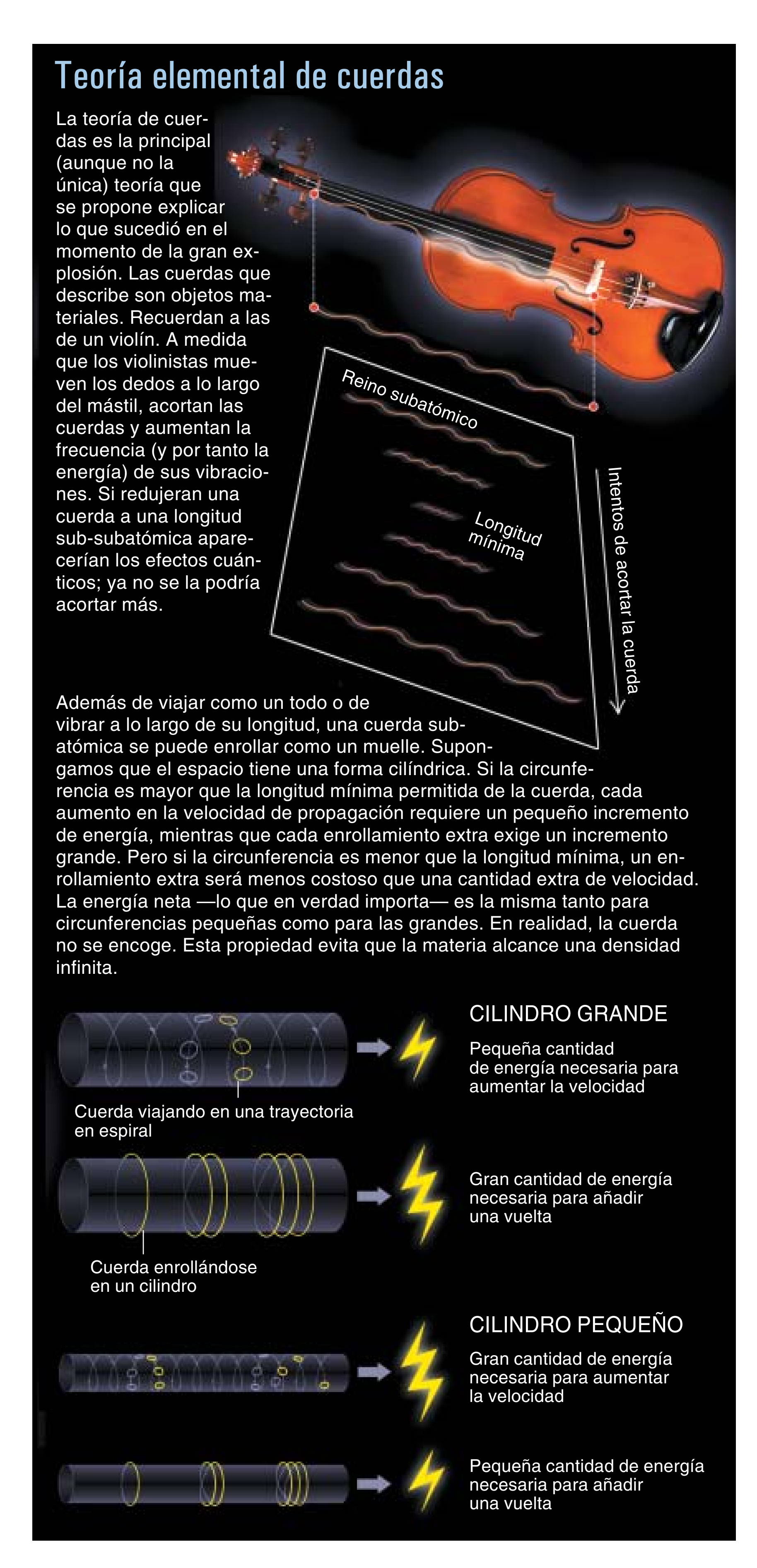
cesitaban nuevas técnicas de cálculo, sino las profundas modificaciones aportadas por la teoría electrodébil de Sheldon Glashow, Steven Weinberg y Abdus Salam a finales de los años sesenta.

La segunda posibilidad, que considero más prometedora, es la teoría de cuerdas, una modificación verdaderamente revolucionaria de la teoría de Einstein. Este artículo se referirá a ella, aunque los proponentes de la gravedad cuántica de bucles afirman que están alcanzando muchas de las mismas conclusiones.

La teoría de cuerdas surgió a partir de un modelo que planteé en 1968 para describir el mundo de las partículas nucleares (protones y neutrones) y sus interacciones. Aunque apasionó en un principio, el modelo fracasó. Sería abandonado años después en favor de la cromodinámica cuántica, que describe las partículas nucleares en términos de constituyentes más elementales, los quarks. Los quarks quedan confinados dentro de los protones o neutrones como si estuvieran ligados entre sí por cuerdas elásticas. Mirando hacia atrás, se comprende que la teoría de cuerdas original captó ese aspecto del mundo nuclear. Sólo más tarde reviviría en el empeño por combinar la relatividad general y la teoría cuántica.

La idea básica es que las partículas elementales no son puntos, sino objetos unidimensionales infinitamente delgados, las cuerdas. El gran zoo de las partículas elementales, cada una con sus propiedades características, refleja los muchos posibles modos de vibración de una cuerda. ¿Cómo puede una teoría tan simple describir el complicado mundo de las partículas y sus interacciones? La respuesta puede encontrarse en lo que cabe llamar magia de la cuerda cuántica. Una vez que las reglas de la mecánica cuántica se aplican a una cuerda vibrante —como si fuera una cuerda de violín en miniatura, excepto que las vibraciones se propagan por ella a la velocidad de la luz—, aparecen nuevas propiedades. Todas tienen hondas consecuencias para la física de partículas y la cosmología.

En primer lugar, las cuerdas cuánticas tienen un tamaño finito. Si no fuera por los efectos cuánticos, se podría cortar una cuerda de violín por la mitad, cortarla de nuevo y así sucesivamente, hasta convertirla en una partícula puntual, sin masa. Pero el principio de incertidumbre de Heisenberg finalmente hace su aparición y evita que las



cuerdas más ligeras se corten en trozos menores que unos 10^{-34} metros. Este irreducible cuanto de longitud, l_s, es una nueva constante de la naturaleza introducida por la teoría de cuerdas, paralelamente a la velocidad de la luz, c, y la constante de Planck, h. Desempeña un papel crucial en casi todos los aspectos de la teoría de cuerdas; establece un límite finito a magnitudes que, de otra manera, podrían hacerse cero o infinito.

En segundo lugar, las cuerdas cuánticas pueden tener momento angular incluso aunque carezcan de masa. En la física clásica, el momento angular es una propiedad de un objeto que gira respecto a un eje. La fórmula del momento angular multiplica entre sí velocidad, masa y distancia al eje; por tanto, un objeto sin masa no puede tener momento angular. Pero las fluctuaciones cuánticas cambian la situación. Una pequeña cuerda puede adquirir hasta dos unidades de h de momento angular sin ganar masa. Esta característica es muy bienvenida, ya que se ajusta a las propiedades de los mediadores de todas las fuerzas fundamentales conocidas, así el fotón (para el electromagnetismo) y el gravitón (para la gravedad). Históricamente, fue el momento angular lo que puso sobre la pista de las implicaciones cuántico-gravitatorias de la teoría de cuerdas.

En tercer lugar, las cuerdas cuánticas exigen la existencia de dimensiones extra del espacio, además de las tres usuales. Una cuerda clásica de violín vibrará con independencia de cuáles sean las propiedades del espacio y el tiempo. Una cuerda cuántica es más exigente. Las ecuaciones que describen la vibración se vuelven incoherentes a menos que el espaciotiempo, o bien esté muy curvado (en contradicción con las observaciones), o contenga seis dimensiones espaciales extra.

Y cuarto, las constantes físicas —las de Newton o de Coulomb, por ejemplo—, que aparecen en las ecuaciones de la física y determinan las propiedades de la naturaleza, ya no tienen valores arbitrarios y fijos. Aparecen en la teoría de cuerdas como campos, bastante parecidos al campo electromagnético, que ajustan sus valores diná-

micamente. Pueden haber tenido diferentes valores en diferentes épocas cosmológicas o en remotas regiones del espacio. Incluso hoy en día las "constantes" físicas podrían variar un poco. La observación de cualquier variación proporcionaría un enorme impulso a la teoría de cuerdas.

Uno de esos campos, el dilatón, es la llave maestra de la teoría de cuerdas; determina la intensidad global de todas las interacciones. El dilatón fascina por su valor: se puede reinterpretar como el tamaño de una dimensión extra del espacio, dando un total de 11 dimensiones al espaciotiempo.

Atar los cabos sueltos

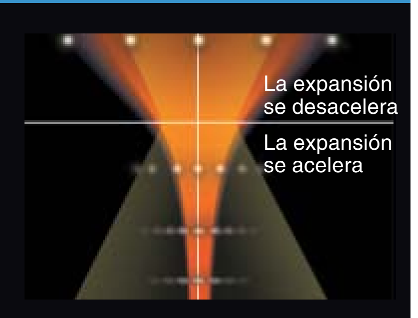
Finalmente, las cuerdas cuánticas han presentado algunas nuevas y sorprendentes simetrías de la naturaleza, las dualidades, que alteran nuestra intuición de qué sucede cuando los objetos se hacen pequeñísimos. Ya he aludido a una forma de dualidad: por lo normal, una cuerda corta es más ligera que una larga, pero si intentamos reducir su tamaño por debajo de la longitud fundamental l_s, se volverá de nuevo pesada.

Otra forma de la simetría, la dualidad T, mantiene que las dimensiones extra pequeñas y grandes son equivalentes. Esta simetría se debe a que las cuerdas pueden moverse de maneras más complicadas que las partículas puntuales. Consideremos una cuerda cerrada (como una sortija) localizada sobre un espacio en forma de cilindro, cuya sección transversal circular represente una dimensión extra finita. Aparte de vibrar, la cuerda puede desplazarse como un todo alrededor del cilindro o enrollarse a su alrededor una o varias veces, como una goma que diese varias vueltas a un tubo (*véase el recuadro* "Teoría elemental de cuerdas").

El coste energético de estos dos estados de la cuerda depende del tamaño del cilindro. La energía del enrollamiento es directamente proporcional al radio del cilindro: los cilindros mayores requieren que la cuerda se estire más para enrollarse a su alrededor; por tanto, el enrollamiento contendrá más energía que la que contendría en un cilindro más pequeño. La energía asociada con el movimiento en torno al círculo, por otro lado, es inversamente proporcional al radio: los cilindros mayores permiten longitudes de onda más largas (menores frecuencias), lo que representa menos energía que las longitudes de onda cortas. Si se sustituye el cilindro grande por otro pequeño, los dos estados de movimiento pueden permutar sus papeles. Las energías que había producido el movimiento circular, se producirán ahora por los

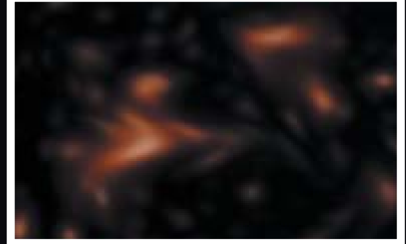
LA HIPOTESIS DE LA EPOCA PREVIA A LA GRAN EXPLOSION

UN INTENTO PRIMERO de aplicar la teoría de cuerdas a la cosmología fue la hipótesis de la pre-gran explosión, según la cual la explosión no es el origen último del universo, sino una transición. Antes, la expansión se aceleraba; después se desaceleró (por lo menos en un principio). La trayectoria de una galaxia por el espaciotiempo (*derecha*) tendría la forma de una copa de vino.





El universo ha existido desde siempre. En el pasado remoto, estaba casi vacío. Las fuerzas, como la gravitación, eran débiles de por sí.



Las fuerzas se fueron intensificando; la materia empezó a acumularse. En algunas regiones se hizo tan densa, que se formó un agujero negro.



Dentro del agujero, el espacio se expandió a un ritmo acelerado. La materia del interior se desconectó de la materia exterior.



Dentro del agujero la materia cayó hacia el centro y aumentó su densidad hasta que alcanzó el límite impuesto por la teoría de cuerdas.



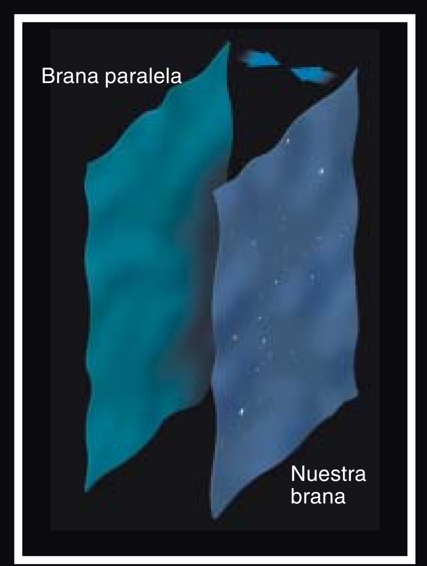
Cuando la materia alcanzó la densidad máxima permitida, los efectos cuánticos provocaron que rebotara en una gran explosión. Fuera, se empezaron a formar otros agujeros, cada uno de ellos un universo distinto.

SAMUEL VELASCO

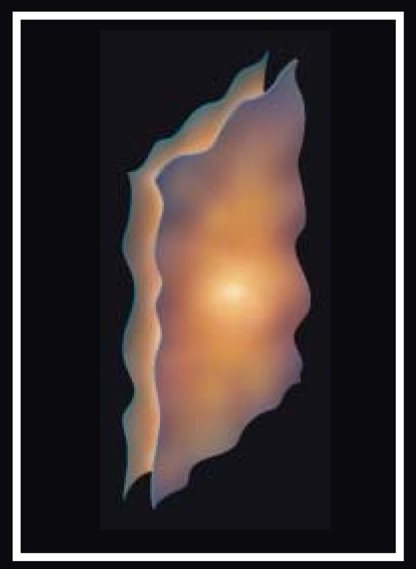
EL MODELO ECPIROTICO

Si nuestro universo es una membrana multidimensional, o simplemente una "brana", moviéndose por un espacio de más dimensiones, la gran explosión podría haber sido la colisión de nuestra brana con una brana paralela. Las colisiones podrían ocurrir cíclicamente. Cada galaxia sigue una trayectoria con forma de reloj de arena a través del espaciotiempo (*abajo*).





Dos branas casi vacías se atraen entre sí. Cada una se está contrayendo en una dirección perpendicular a su movimiento.



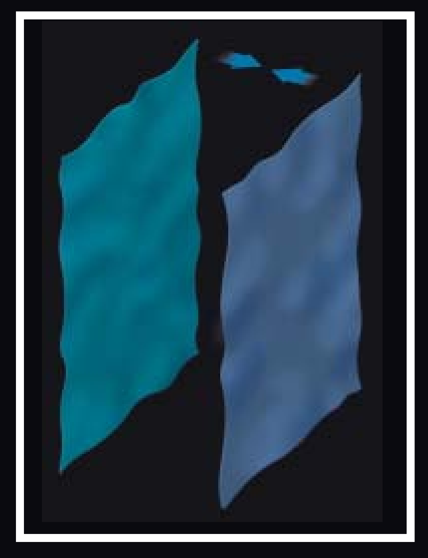
Las branas chocan, convirtiendo sus energías cinéticas en materia y radiación. Esta colisión es la gran explosión.



Las branas rebotan. Empiezan a expandirse a un ritmo desacelerado. La materia se agrupa en estructuras tales como los cúmulos de galaxias.



En el modelo cíclico, a medida que las branas se separan, la fuerza atractiva entre ellas las frena. La materia se reduce.



Las branas acaban de separarse y empiezan a acercarse. Durante la inversión, cada brana se expande a un ritmo acelerado.

SAMUEL VELASCO

enrollamientos, y viceversa. Un observador exterior sólo nota los niveles de energía, no el origen de dichos niveles. Para tal observador, los radios grandes y los pequeños son físicamente equivalentes.

Aunque la dualidad T se suele formular para espacios cilíndricos, en los que una dimensión (la circunferencia) es finita, hay una variante que se aplica a nuestras tres dimensiones corrientes, que nos parece que cabe estirar indefinidamente. Se debe ser cuidadoso al hablar de la expansión de un espacio infinito. Su tamaño global no puede cambiar; sigue siendo infinito. Pero sí se puede expandir en el sentido de que los cuerpos de su interior —las galaxias— se alejen unos de otros. La variable crucial no es el tamaño del espacio como un todo, sino su factor de escala, el factor por el que la distancia entre galaxias cambia; se manifiesta en el desplazamiento hacia el rojo de la luz galáctica que los astrónomos observan. Según la dualidad T, universos con pequeños factores de escala son equivalentes a los de

factores de escala grandes. Una tal simetría no aparece en las ecuaciones de Einstein. Surge de la unificación, implícita en la teoría de cuerdas; al dilatón le toca ahí un papel central.

Durante años, los teóricos de las cuerdas pensaron que la dualidad T se aplicaba sólo a las cuerdas cerradas y no a las cuerdas abiertas, que tienen cabos sueltos y, por tanto, no se pueden enrollar. En 1995 Joseph Polchinski, de la Universidad de California en Santa Barbara, observó que la dualidad T se aplica también a las cuerdas abiertas, siempre que el cambio entre radios grandes y pequeños se acompañe de un cambio en las condiciones en los puntos finales de la cuerda. Hasta entonces, se habían dictado condiciones de contorno en las que ninguna fuerza actuaba en los extremos de las cuerdas, dejándolos libres para moverse. Bajo la dualidad T, estas condiciones se convierten en las llamadas condiciones de contorno de Dirichlet y los extremos quedan fijos.

Cualquier cuerda dada puede mezclar ambos tipos de condiciones de contorno. Por ejemplo, los electrones pueden ser cuerdas cuyos extremos se mueven libremente en tres de las 10 dimensiones espaciales, pero están fijos en las otras siete. Aquellas tres dimensiones forman un subespacio conocido como membrana de Dirichlet, o D-brana. En 1996 Petr Horava, de la Universidad de California en Berkeley, y Edward Witten, del Instituto de Estudios Avanzados de Princeton, propusieron que nuestro universo reside en una tal brana. La movilidad parcial de los electrones y de otras partículas explica por qué somos incapaces de percibir las diez dimensiones del espacio en todo su esplendor.

Domesticar el infinito

Las mágicas propiedades de las cuerdas cuánticas apuntan a una misma conclusión: las cuerdas aborrecen el infinito. Como no pueden contraerse hasta convertirse en un punto infinitesimal, evitan las paradojas que un colapso así entraña. Su tamaño no nulo y nuevas simetrías establecen límites superiores a las magnitudes físicas que aumentan ilimitadamente en las teorías ordinarias, e imponen cotas inferiores a las que decrecen. Los teóricos de cuerdas piensan que cuando se proyecta la historia del universo hacia atrás en el tiempo, la curvatura del espaciotiempo va aumentando. Pero en vez de ir creciendo hasta el infinito (hasta la tradicional singularidad de la gran explosión), acaba por alcanzar un máximo y se reduce otra vez. Antes de la teoría de cuerdas, difícil era imaginar mecanismos que eliminasen tan limpiamente la singularidad.

Las condiciones cerca del tiempo cero de la gran explosión eran tan extremas, que aún no se sabe cómo resolver las ecuaciones. Sin embargo, los teóricos de cuerdas han aventurado hipótesis acerca del universo anterior a la gran explosión. Están tomando en consideración dos modelos sobre todo.

El primero, que establece la existencia de una época previa a la gran explosión —mis compañeros y yo empezamos a desarrollarlo en 1991—, combina la dualidad T con la simetría, más conocida por todos, de la inversión temporal, según la cual las ecuaciones de la física valen tanto cuando se las aplica hacia atrás en el tiempo como cuando se las aplica hacia delante. La combinación da lugar a nuevas cosmologías posibles en las que el universo se expande, digamos que cinco segundos antes de la gran explosión, a la misma velocidad que cinco segundos después. Pero el cambio de la expansión era opuesto en los dos instantes; si se estaba desacelerando después, antes se aceleraba. En pocas palabras: la gran explosión podría no haber sido el origen del universo, sino sólo una violenta transición desde la aceleración a la desaceleración.

La belleza de este modelo es que incorpora de manera automática la gran intuición de la teoría inflacionaria estándar: que el universo debió sufrir un período de aceleración a fin de quedar tan homogéneo e isótropo. En la teoría estándar, la aceleración ocurre después de la gran explosión debido a un campo, el inflatón, *ad hoc*. En el modelo de la pre-gran explosión, ocurre antes de la explosión, como un resultado natural de las nuevas simetrías de la teoría de cuerdas.

Según esa hipótesis, el universo pre-explosivo era casi una imagen especular perfecta del que siguió a la explosión (*véase el recuadro* “La hipótesis de la época previa a la gran explosión”). Si el universo es eterno hacia el futuro, con su contenido enrareciéndose hasta convertirse en un caldo diluido, también será eterno hacia el pasado. Infinitamente antes, estaba casi vacío; sólo contenía un gas tenue, muy disperso y caótico, de radiación y materia. Las fuerzas de la naturaleza, controladas por el campo del dilatón, eran tan débiles que las partículas del gas, apenas si interaccionaban.

A medida que transcurrió el tiempo, las fuerzas ganaron intensidad y tendieron a agrupar la materia. Aleatoriamente, algunas regiones acumularon materia a expensas de sus entornos. Por último, la densidad en estas regiones se hizo tan alta, que empezaron a formarse agujeros negros. La materia del interior de esas regiones se desconectó entonces del exterior. El universo se partió en fragmentos disconexos.

Dentro de un agujero negro, el espacio y el tiempo intercambian papeles. El centro del agujero negro no es un punto del espacio sino un instante de tiempo. A medida que la materia se precipitaba hacia el centro, alcanzaba densidades más y más altas. Pero cuando la densidad, la temperatura y la curvatura adquirieron los valores máximos permitidos por la teoría de cuerdas, rebotaron y empezaron a decrecer. El momento de esta inversión es lo que llamamos gran explosión. El interior de uno de esos agujeros negros se convirtió en nuestro universo.

No es sorprendente que un esquema tan poco corriente haya provocado controversias. Andrei Linde, de la Universidad de Stanford, ha argumentado que para que concordase con las observaciones, el agujero negro del que surgió nuestro universo tendría que haberse formado con un tamaño inusualmente grande, mucho mayor que la escala de longitud de la teoría de cuerdas. Una respuesta a esta objeción es que las ecuaciones predicen agujeros negros de todos los tamaños posibles. Resulta que nuestro universo

OBSERVACIONES

LA OBSERVACION del universo anterior a la gran explosión parecerá una tarea imposible, pero puede que sobreviva de esa época una forma de radiación: la radiación gravitatoria. Estas variaciones periódicas del campo gravitatorio se detectarían indirectamente por sus efectos sobre la polarización del fondo cósmico de microondas (*vista simulada inferior*) o bien directamente en observatorios terrestres. Tanto la hipótesis de la época pre-gran explosión como la ecpirótica predicen más ondas gravitatorias de alta frecuencia y menos de baja frecuencia que los modelos inflacionarios ordinarios (*abajo*). Las medidas existentes de varios fenómenos astronómicos no permiten distinguir entre estos modelos, pero las futuras observaciones del satélite Planck y de los observatorios LIGO y VIRGO deberían poder hacerlo.

Densidad de energía de las ondas gravitatorias (unidades cosmológicas)
1
10^{-5}
10^{-10}
10^{-15}
10^{-20}
Precisión de los nuevos LIGO/VIRGO
Modelo de la pre-gran explosión
Inflación ordinaria
Precisión del satélite Planck
Modelo ecpirótico
10^{-15} 10^{-10} 10^{-5} 1 10^{5} 10^{10}
Frecuencia de las ondas gravitatorias (hertz)

se formó dentro de uno suficientemente grande.

Una objeción más seria, planteada por Thibault Damour, del Instituto de Altos Estudios Científicos, de Bures-sur-Yvette, y Marc Henneaux, de la Universidad Libre de Bruselas, es que la materia y el espaciotiempo se habrían comportado caóticamente cerca del momento de la explosión, en posible contradicción con la observada regularidad del universo primitivo. Hace poco he propuesto que un estado caótico produciría un denso gas de "agujeros de cuerda" en miniatura, unas cuerdas tan pequeñas y con tanta masa, que estarían al borde de convertirse en agujeros negros. El comportamiento de estos agujeros podría resolver el problema planteado por Damour y Henneaux. Tom Banks, de la Universidad de Rutgers, y Willy Fischler, de la de Texas en Austin, han ofrecido una solución semejante. También existen otras críticas; está por determinar si han descubierto un fallo fatal de la hipótesis.

Choques de branas

El otro modelo importante de un universo anterior a la explosión es el ecpirótico. [O "a partir del fuego"; se refiere a la creencia de la filosofía estoica en el final y reconstrucción cíclicos del mundo por el fuego.] Lo concibió hace tres años un equipo de cosmólogos y teóricos de cuerdas —Justin Khoury, de la Universidad de Columbia, Paul Steinhardt, de Princeton, Burt A. Ovrut, de la de Pennsylvania, Nathan Seiberg, del Instituto de Estudios Avanzados, y Neil Turok, de la Universidad de Cambridge—. Se basa en la hipótesis de que nuestro universo es una de muchas D-branas que flotan en un espacio de más dimensiones. Las branas ejercen fuerzas atractivas entre ellas y, ocasionalmente, chocan. La gran explosión consistiría en el impacto de otra brana sobre la nuestra (*véase el recuadro* "El modelo ecpirótico").

En una variante de este modelo, las colisiones ocurren cíclicamente. Dos branas chocan, rebotan, se separan, se atraen mutuamente, chocan de nuevo, y así sucesivamente. Entre colisiones, las branas son plásticas: se expanden a medida que se alejan y se contraen algo a medida que se vuelven a acercar. Durante la transición, el ritmo de expansión se acelera; la actual expansión acelerada del universo quizás augure una nueva colisión.

Los modelos de la pre-gran explosión y ecpirótico poseen algunas características comunes. Ambos empiezan con un universo grande, frío y casi vacío, y ambos comparten el difícil (y no resuelto) problema de la transición entre la fase previa a la explosión y lo que viene después. Matemáticamente, la principal diferencia entre ambas hipótesis se encuentra en el comportamiento del campo del dilatón. En la pregran explosión, el dilatón empieza con un valor pequeño —de manera que las fuerzas de la naturaleza son débiles— y paulatinamente gana intensidad. Lo opuesto sucede en el modelo ecpirótico, donde la colisión ocurre cuando las fuerzas más débiles son.

Los creadores de la teoría ecpirótica esperaban que la debilidad de las fuerzas permitiría analizar el rebote con más facilidad, pero hubieron de enfrentarse de todas formas a las dificultades de las curvaturas grandes. Está por ver si su modelo realmente evita una singularidad. Además, requiere unas muy especiales condiciones para resolver las usuales paradojas cosmológicas. Por ejemplo, las branas a punto de chocar debieron haber sido casi exactamente paralelas entre sí, pues de lo contrario la colisión no podría haber dado lugar a una explosión suficientemente homogénea. La versión cíclica podría resolver este problema, ya que las sucesivas coli-

NASA (*arriba*); GABRIELE VENEZIANO CERN (*abajo*)

siones permitirían que las branas se rectificaran por sí mismas.

Dejando aparte la difícil tarea de justificar plenamente estos dos modelos desde el aspecto matemático, hay que preguntarse si tienen algunas consecuencias físicas observables. A primera vista, ambos podrían parecer un ejercicio, no de física, sino de metafísica, ideas interesantes de las que los observadores nunca podrán probar si son ciertas o no. Tal actitud es demasiado pesimista. Como sucede con los detalles de la fase inflacionaria, los de una posible época pre-explosiva podrían tener consecuencias observables, inscritas sobre todo en las pequeñas variaciones observadas en la temperatura del fondo cósmico de microondas.

En primer lugar, las observaciones muestran que las fluctuaciones de temperatura fueron moldeadas por ondas acústicas durante varios cientos de miles de años. La regularidad de las fluctuaciones indica que las ondas estuvieron sincronizadas. Se han descartado muchos modelos cosmológicos a lo largo de los años porque no podían explicar esta sincronía. Los modelos inflacionario, de la pre-gran explosión y ecpirótico superan esta primera prueba. En los tres, las ondas fueron generadas por procesos cuánticos amplificados durante el período de expansión cósmica acelerada. Las fases de las ondas estaban alineadas.

En segundo lugar, cada modelo predice una diferente distribución de las fluctuaciones de la temperatura con respecto al tamaño angular. La observación ha encontrado que las fluctuaciones de todos los tamaños tienen aproximadamente la misma amplitud. (Las desviaciones discernibles tienen lugar sólo a escalas muy pequeñas, donde las fluctuaciones primordiales han sido alteradas por procesos subsiguientes.) Los modelos inflacionarios reproducen muy bien esta distribución. Durante la inflación, la curvatura del espacio cambió de un modo paulatino; se generaron, pues, fluctuaciones de diferentes tamaños bajo condiciones muy parecidas. En los dos modelos basados en la teoría de cuerdas, la curvatura evolucionó rápidamente, con lo que creció la amplitud de las fluctuaciones de pequeña escala, pero otros procesos potenciaron las de escalas grandes: todas las fluctuaciones quedaron con la misma intensidad. En el modelo ecpirótico, entre esos otros procesos se incluye la dimensión extra del espacio, la que separa las branas que chocan; en el de la pre-gran explosión, un campo cuántico relacionado con el dilatón, el axión. En resumidas cuentas, los tres modelos se ajustan a los datos.

En tercer lugar, las variaciones de temperatura pueden deberse a dos procesos distintos del universo primitivo: las fluctuaciones de la densidad de materia y las ondulaciones causadas por las ondas gravitatorias. En la inflación intervienen ambos procesos, mientras que en la pre-gran explosión y el modelo ecpirótico predominan las variaciones de densidad. Las ondas gravitatorias de ciertos tamaños dejarían una señal distintiva en la polarización del fondo de microondas [*véase* "Ecos de la Gran Explosión", por Robert R. Caldwell y Marc Kamionkowski; INVESTIGACIÓN Y CIENCIA, marzo 2001]. Futuros observatorios, como el satélite Planck de la Agencia Europea del Espacio, captarán esa señal, si existe; se tratará de una contrastación casi definitiva.

Un cuarto examen se refiere a la estadística de las fluctuaciones. En la inflación, las fluctuaciones siguen una curva con forma de campana, o gaussiana. Lo mismo puede suceder en el caso ecpirótico, mientras que la pre-gran explosión permite apreciables desviaciones de la curva de Gauss.

Los análisis del fondo de microondas no son la única manera de comprobar estas teorías. Con pre-gran explosión debería haber también un fondo aleatorio de ondas gravitatorias en un intervalo de frecuencias que, aunque irrelevante para el fondo de microondas, sería detectable por los futuros observatorios de ondas gravitatorias. Además, debido a que los modelos de la pre-gran explosión y ecpirótico implican cambios en el campo del dilatón, que va acoplado al campo electromagnético, ambos llevarían a fluctuaciones a gran escala del campo magnético; podrían manifestarse vestigios suyos en los campos magnéticos galácticos e intergalácticos.

¿Cuándo empezó el tiempo? Aún no tenemos una respuesta concluyente, pero por lo menos dos teorías potencialmente comprobables mantienen con verosimilitud que el universo —y por tanto el tiempo— existía antes de la gran explosión. Si una u otra son ciertas, el cosmos habría existido siempre, y aunque un día vuelva a derrumbarse sobre sí mismo, no acabará nunca.

El autor

Gabriele Veneziano, físico teórico del CERN, fue el padre de la teoría de cuerdas a finales de los años sesenta; por eso ha recibido este año el premio Heineman, de la Sociedad Americana de Física y del Instituto de Física. En aquel tiempo, se consideró que la teoría fallaba. No lograba su objetivo de explicar el núcleo atómico; Veneziano pronto desvió su atención a la cromodinámica cuántica, a la que hizo importantes contribuciones. Después de que la teoría de cuerdas reapareciese como una teoría de la gravedad en los años ochenta, sería uno de los primeros que la aplicaron a los agujeros negros y la cosmología.

Bibliografía complementaria

SUPERSTRING COSMOLOGY. James E. Lidsey, David Wands y Edmund J. Copeland en *Physics Reports*, vol. 337, n.os 4-5, págs. 343-492; octubre 2000.

FROM BIG CRUNCH TO BIG BANG. Justin Khoury, Burt A. Ovrut, Nathan Seiberg, Paul J. Steinhardt y Neil Turok en *Physical Review D*, vol. 65, n.º 8, Trabajo no.086007; 15 de abril, 2002.

A CYCLIC MODEL OF THE UNIVERSE. Paul J. Steinhardt y Neil Turok en *Science*, vol. 296, n.º 5572, páginas 1436-1439; 24 de mayo, 2002.

EL UNIVERSO ELEGANTE. Brian Greene. Ed. Crítica; Barcelona, 2003.

THE PRE-BIG BANG SCENARIO IN STRING COSMOLOGY. Maurizio Gasperini y Gabriele Veneziano en *Physics Reports*, vol. 373, n.os 1-2, páginas 1-212; enero 2003.

MicroARN

El descubrimiento de unas moléculas de ARN diminutas en las células eucariotas ha modificado de raíz nuestro conocimiento sobre los mecanismos de regulación de la expresión génica

César Llave

A principio de los años noventa del siglo pasado, el estudio genético del nemátodo *Caenorhabditis elegans* revelaba la existencia de unas moléculas de ácido ribonucleico (ARN) diminutas y peculiares. Se las denominó pequeños ARN transitorios, pues se hallaban asociados con el control temporal del desarrollo larvario.

La investigación ulterior descubrió que esos pequeños ARN eran, en verdad, reguladores de la expresión de ciertos genes. Inhibían la traducción de sus ARN mensajeros en momentos clave del desarrollo. (El ARNm porta la información necesaria para la síntesis de proteína.) Desde entonces, y fruto del interés creciente hacia el conocimiento de tales moléculas, se han identificado numerosos ARN de pequeño tamaño con capacidad de modular la expresión génica en organismos muy dispares: protozoos, hongos, plantas y animales.

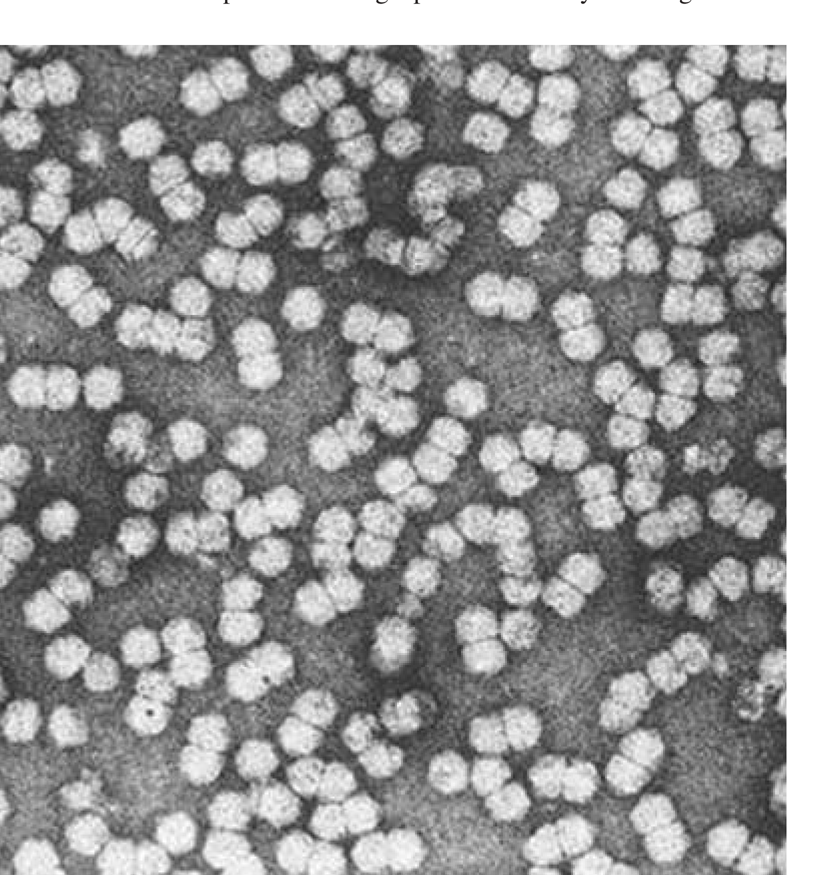
Los pequeños ARN transitorios, o temporales, forman parte de un grupo numeroso y heterogéneo de ácidos ribonucleicos que no codifican proteínas. Se trata de los microARN. Aunque comenzamos a conocer sus funciones, queda mucho por averiguar en torno a su misión reguladora de genes y sus mecanismos de operación. Poco antes de que se descubrieran los microARN, se dio con un proceso de inesperado alcance: la interferencia por ARN. De este mecanismo de silenciamiento génico subsiguiente a la transcripción se encarga un segundo tipo de moléculas de ARN, los pequeños ARN interferentes.

ARN interferente

El silenciamiento por ARN constituye un proceso altamente específico de degradación del ácido ribonucleico. Se desarrolla después de la transcripción génica. Afecta, pues, a la viabilidad de los ARN producidos por los genes. Para poner en marcha ese mecanismo degradativo de desactivación génica se requiere la presencia en el citoplasma de moléculas de ARN bicatenario con una secuencia de nucleótidos idéntica a la de los genes diana.

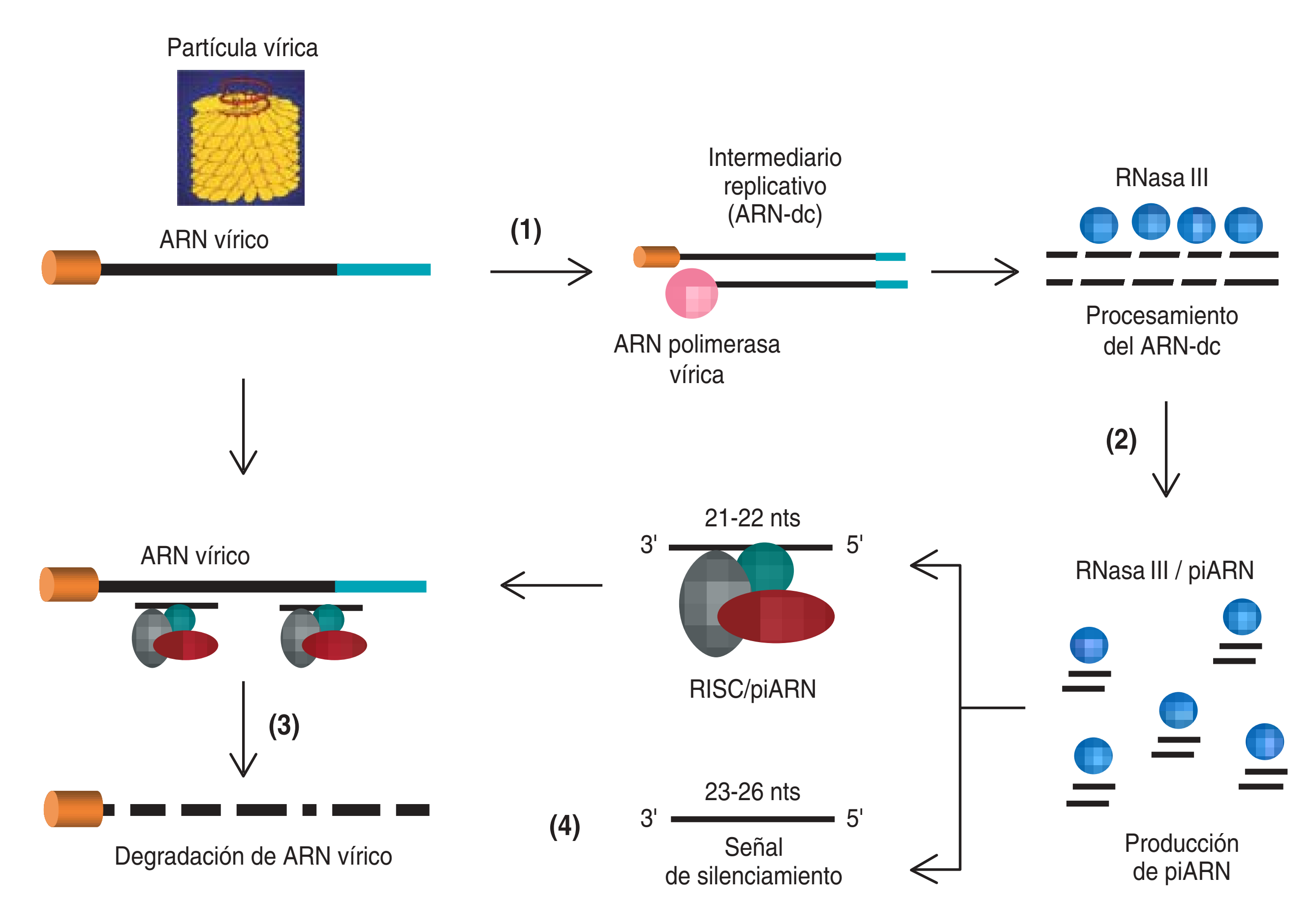
Los ARN bicatenarios son reconocidos por una RNasa de tipo III, específica de ARN de doble cadena. Esta enzima los procesa para dar lugar a pequeños dúplices de ARN, de 21 a 26 nucleótidos de longitud. De entre estos diminutos ARN interferentes, los menores (21 a 23 nucleótidos) se integran en un complejo ribonucleoproteico (RISC), donde actúan a modo de guías para el reconocimiento de los ARN diana. Fruto de la interacción entre este complejo y el ARN mensajero homólogo se produce la degradación y desactivación de este último.

Además, los pequeños ARN interferentes pueden unirse por sí mismos con los ARN diana; operan entonces como iniciadores en una reacción de polimerización catalizada por una ARN polimerasa dependiente de ARN. Utilizando de molde el ARN diana, se genera un ARN bicatenario, que es reconocido de nuevo por la RNasa III y degradado ("digerido"), en pequeños ARN interfe-



1. Fotografía al microscopio electrónico de partículas víricas del virus del mosaico dorado de la judía (*Begomovirus germiniviridae*).

CESAR LLAVE; *FUENTE*: COMITE INTERNACIONAL DE TAXONOMIA DE VIRUS



2. MODELO DE SILENCIAMIENTO GENICO en las plantas, inducido por un virus. Una vez en el citoplasma celular, las partículas víricas se desensamblan y liberan su ARN genómico. El virus utiliza sus propias ARN polimerasas para replicarse en la célula infectada; se vale para ello de intermediarios de ARN de doble cadena (ARN-dc) (*1*). Estos inductores de silenciamiento se reconocen y procesan por una RNasa III, para originar pequeños ARN interferentes (piARN) (*2*). Los piARN de menor tamaño se asocian con un complejo enzimático (RISC) para el reconocimiento y posterior degradación del ARN homólogo vírico (*3*). Los piARN de mayor tamaño se hallan asociados con la señal sistémica, que lleva la respuesta de silenciamiento a otras partes de la planta (*4*).

CESAR LLAVE

rentes. En numerosos experimentos se ha confirmado que basta la sola expresión "artificial" de moléculas inductoras (ARN bicatenario) en tejidos vegetales o en cultivos celulares para disparar la respuesta de silenciamiento, incluso en ausencia de posibles genes diana.

Los pequeños ARN interferentes de mayor tamaño (24 a 26 nucleótidos) no participan en la degradación de los ARN mensajeros. Antes bien, se desplazan a otras células del organismo, donde inician una nueva respuesta de silenciamiento génico.

Silenciamiento génico en eucariotas

Pero, ¿qué funciones cumple el silenciamiento génico subsiguiente a la transcripción en los organismos eucariotas? Se ha comprobado que tal represión constituye un mecanismo de defensa contra ácidos nucleicos exógenos (transposones, por ejemplo) y, en el caso de las plantas, contra los virus. Cerca del 90 % de los virus vegetales poseen genomas de ARN de cadena sencilla, que se multiplican en la célula infectada a través de ARN bicatenarios muy estables.

Estos virus replicativos son inductores muy potentes que disparan el mecanismo de interferencia por ARN y permiten que la célula infectada degrade el ARN vírico, para así superar la infección. El grupo de David C. Baulcombe, del Laboratorio Sainsbury en Norwich, Inglaterra, detectó pequeños ARN interferentes correspondientes a secuencias de varios virus vegetales (así, el virus X de la patata o el virus del cascabeleo del tabaco) en plantas infectadas. Demostró que los virus eran reconocidos como objetivo a abatir por la maquinaria celular de silenciamiento.

El grupo de Baulcombe demostró también que la infección de plantas transgénicas conducía a la represión del transgén; se trataba de plantas que expresaban constitutivamente el gen de la proteína de fluorescencia verde (GFP) infectadas con un virus quimera que portaba este mismo gen marcador integrado en su genoma. Si la expresión del genoma vírico está regulada post-transcripcionalmente por un mecanismo de silenciamiento, también lo estará la expresión de cualquier secuencia homóloga a la vírica; en este caso, GFP formaba parte de la secuencia vírica en el virus quimera.

Pequeños ARN endógenos

La producción y acumulación de pequeños ARN interferentes constituye, pues, un paso clave en la ruta de silenciamiento génico. Su síntesis se produce en repuesta a un inductor exógeno de silenciamiento, sea la forma replicativa de un virus o sea cualquier otro ARN con



3. Microfotografía de partículas víricas del virus "y" de la patata (*Potyvirus potyviridae*).

estructura bicatenaria. Pero de la misma manera que la maquinaria celular está preparada para regular la expresión de genes foráneos por mediación de pequeños ARN interferentes, los organismos eucariotas han desarrollado diversas estrategias, algunas de ellas basadas en la codificación y síntesis de pequeños ARN, para el control de la expresión de sus propios genes.

En diversos laboratorios se ha venido aplicando, casi simultáneamente, una misma estrategia para la multiplicación ("amplificación"), clonación y secuenciación de pequeños ARN endógenos en protozoos, plantas, nemátodos, moscas y mamíferos. La disponibilidad del genoma completo del nemátodo *C. elegans*, de *Drosophila melanogaster* y de la planta *Arabidopsis thaliana* ha facilitado la caracterización de estas diminutas moléculas de ARN endógenas. En particular, nos ha permitido acotar el origen de cada pequeño ARN en el genoma de estos organismos, identificar entre ellos microARN, pequeños ARN interferentes, así como otras muchas moléculas de diminutos ARN cuya estructura, biogénesis y función nos son todavía desconocidas. Además, el estudio de todos estos ARN nos ha permitido descubrir genes potencialmente regulados por un fenómeno de silenciamiento génico post-transcripcional e identificar posibles genes sujetos a una regulación mediada por microARN.

En el caso de *Arabidopsis* se han definido más de 200 secuencias únicas de pequeños ARN. Algunas de estas secuencias presentan homología con una sola región del genoma, mientras que otras se alojan en múltiples loci dentro de los cinco cromosomas de la planta. Su longitud es de 21 a 25 nucleótidos, similar a los pequeños ARN interferentes originados en respuesta a un fenómeno de silenciamiento inducido por un ARN bicatenario. Estos pequeños ARN presentan homología perfecta de secuencia con genes codificadores de proteínas, transposones o genes estructurales, aunque la mayoría son homólogos con regiones intergénicas del genoma.

Funciones de los pequeños ARN endógenos

Cerca del 20 % de los pequeños ARN descritos en plantas presentan homología perfecta con secuencias genómicas dentro de regiones codificadoras de proteínas, pertenezcan a genes o a transposones. Basados en ese fenómeno peculiar de los vegetales, cabe sospechar que esta clase de pequeños ARN pudiera ser el resultado del silenciamiento de ARN endógenos.

En el caso de los transposones, existen pruebas que ponen de manifiesto la capacidad de las plantas para limitar su expresión por medio del silenciamiento génico. De hecho, la naturaleza altamente repetitiva de los transposones, unida a la alta probabilidad de que se integren en el genoma y originen por ende transcritos con el potencial de formar estructuras bicatenarias, facilita que sean dianas de silenciamiento post-transcripcional. Habida cuenta del número estimado de transposones en el genoma de *Arabidopsis*, el porcentaje de pequeños ARN localizados en transposones es sustancialmente mayor que el de alojados en genes codificadores de proteínas. De ello se desprende que tales moléculas intervienen de una manera decisiva en el control de la expresión de estos elementos móviles.

De forma análoga, la expresión de genes codificadores de proteínas podría estar modulada por una estrategia reguladora similar. Algunos de los genes que portan secuencias homólogas con pequeños ARN producen preARN mensajeros cuya configuración en dúplex estable podría ser reconocida por la RNasa III específica de doble cadena para su procesamiento en pequeños ARN interferentes.

En otros casos, los ARN mensajeros diana podrían servir de molde para la síntesis de un ARN bicatenario en una reacción catalizada por un ARN polimerasa dependiente de ADN; éste, lo mismo que el anterior, sufriría una fragmentación, por parte de la maquinaria celular de silenciamiento, en pequeños ARN interferentes que contribuirían a la degradación de secuencias homólogas o estrechamente emparentadas.

En conclusión, puesto que la expresión de transposones y algunos genes codificadores de proteínas parece estar regulada por un proceso de silenciamiento génico, cabe pensar que una porción de los pequeños ARN existentes en el citoplasma celular forman parte de una población más extensa de pequeños ARN interferentes. La planta puede, así, recurrir a esta herramienta molecular, no sólo para defenderse ante los virus, sino también para controlar la expresión de sus propios genes.

Regiones intergénicas

Aproximadamente el 90 % de los pequeños ARN encontrados en *Arabidopsis thaliana* y en otros organismos se corresponde con regiones intergénicas. Se sabe que tales zonas desempeñan un papel crucial en la producción de microARN implicados en la regulación de genes endógenos del organismo. Esto es así porque algunos de estos microARN homólogos con regiones intergénicas son al mismo tiempo complementarios con la secuencia de ARN mensajeros de genes endógenos de la planta. Como resultado de dicha complementariedad, los microARN se emparejan con sus ARN mensajeros diana e interfieren con su expresión. Las regiones

CESAR LLAVE

intergénicas no son, en absoluto, meros espaciadores en el curso de la secuencia genómica. Para ser más exactos, algunas de las hasta hoy consideradas regiones intergénicas constituyen, en realidad, genes codificadores de precursores para la síntesis de microARN.

En efecto, las regiones intergénicas que originan estos microARN se corresponden con secuencias dotadas de capacidad suficiente para formar precursores de ARN con estructura secundaria estable. Estos precursores de doble cadena son los sustratos sobre los que actúa una RNasa III, similar a la enzima que da origen a los pequeños ARN interferentes, para generar microARN endógenos.

Regulación de los microARN

Parece demostrado que la acumulación de estos microARN se regula durante el desarrollo del organismo. En el caso de los microARN descritos en *C. elegans*, su producción tiene lugar a lo largo de las distintas fases del desarrollo del gusano. Tampoco en las plantas encontramos una síntesis aleatoria de microARN, sino que, en su mayoría, los microARN presentan patrones de acumulación específicos para un tejido determinado.

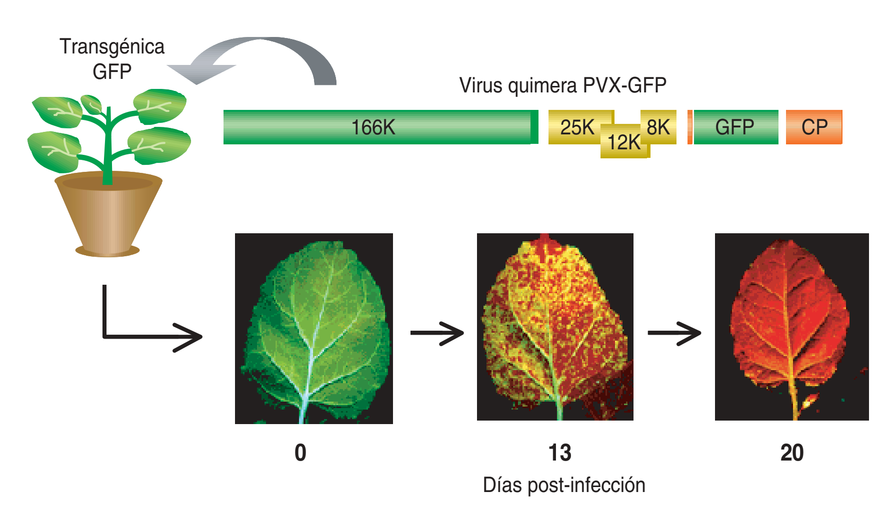
Debe resaltarse también el alto grado de persistencia, a lo largo de la evolución biológica, de la producción de estos microARN. Algunos de los identificados en nemátodos son exactamente iguales que los encontrados en mamíferos; algunos de los microARN encontrados en *A. thaliana* se hallan también presentes en *Nicotiana benthamiana* y en *Oryza sativa*.

MicroARN: biosíntesis y expresión

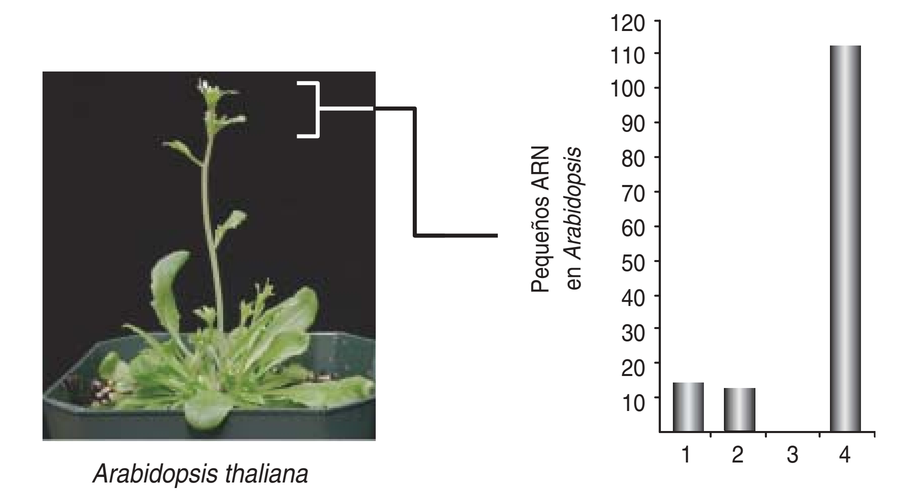
Algunos microARN descritos en plantas y animales presentan características idénticas, en cuanto a su biosíntesis y expresión, a las de los pequeños ARN transitorios de *C. elegans*. De hecho, estos últimos pueden englobarse bajo la denominación común de microARN.

Desde el punto de vista estructural, los microARN son similares a los pequeños ARN interferentes. Presentan, sin embargo, ciertas peculiaridades distintivas. Todos los microARN tienen un tamaño próximo a los 21 nucleótidos y se expresan durante el desarrollo del organismo. Se originan a partir de precursores de ARN con estructura a modo de horquilla estable, que contienen la secuencia del microARN en uno de sus brazos; su producción depende de una enzima RNasa III específica de ARN de doble cadena pero distinta de la que origina los pequeños ARN interferentes.

Si bien todos los microARN responden a unos criterios establecidos de expresión y biogénesis, divergen en su función respectiva, en muchos casos todavía desconocida. Se va avanzando, pese a todo. Los pequeños ARN transitorios codificados por los genes *lin-4* o *let-7* se hallan implicados en el desarrollo temporal del nemátodo *C. elegans* e inhiben la traducción de los ARN mensajeros dianas; interaccionan mediante emparejamiento parcial de bases con su región 3' no codificadora. De esta manera, el reconocimiento del mensajero diana por parte del microARN no depende de



4. LOS VIRUS SON INDUCTORES y diana de la respuesta de silenciamiento génico. Por expresión del transgén, las plantas con la proteína GFP emiten fluorescencia verde bajo iluminación ultravioleta (día 0). Tras la infección con el virus quimera PVX-GFP la hoja se torna roja (coloración de la clorofila bajo luz ultravioleta) debido a la desactivación por silenciamiento génico del transgén GFP (días 13 y 20). PVX es el virus X de la patata.



5. DISTRIBUCION Y LOCALIZACION de pequeños ARN en el genoma de *Arabidopsis thaliana*. En el eje de ordenadas de la gráfica se muestran el número de loci de la planta homólogos con pequeños ARN.

CESAR LLAVE; *FUENTE:* DAVID C. BAULCOMBE, LABORATORIO SAINSBURY, NORWICH INGLATERRA (*arriba*); *FUENTE:* JAMES C. CARRINGTON, UNIVERSIDAD ESTATAL DE OREGON, EE.UU. (*abajo*)

una complementariedad de secuencia absoluta entre ambos, sino que ocurre incluso en presencia de varios errores.

Es posible que algunos de los microARN identificados en animales y plantas, para los que no se han encontrado genes diana con homología de secuencia completa, funcionen como pequeños ARN transitorios; se emparejarían parcialmente con los ARN mensajeros diana y bloquearían su traducción. En el caso de las plantas se han identificado, mediante análisis informático, algunos genes que presentan homología parcial con microARN; podrían, pues, estar regulados potencialmente por un mecanismo de este tipo.

Aparentemente, existen diferencias sustanciales entre los ARN mensajeros regulados por pequeños ARN transitorios y los ARN mensajeros supuestamente modulados por otros microARN. En primer lugar, la secuencia complementaria entre el ARN mensajero de la planta y el microARN se encuentra mayoritariamente en el interior de la región codificadora de la proteína; en la región 3' no codificadora del ARN mensajero se ubica la de los pequeños ARN transitorios *lin-4* y *let-7* en nemátodos. En segundo lugar, la región complementaria entre el ARN mensajero de la planta y el microARN es única, mientras que son múltiples las secuencias complementarias dentro la región 3' no codificadora de los mensajeros regulados por pequeños ARN transitorios.

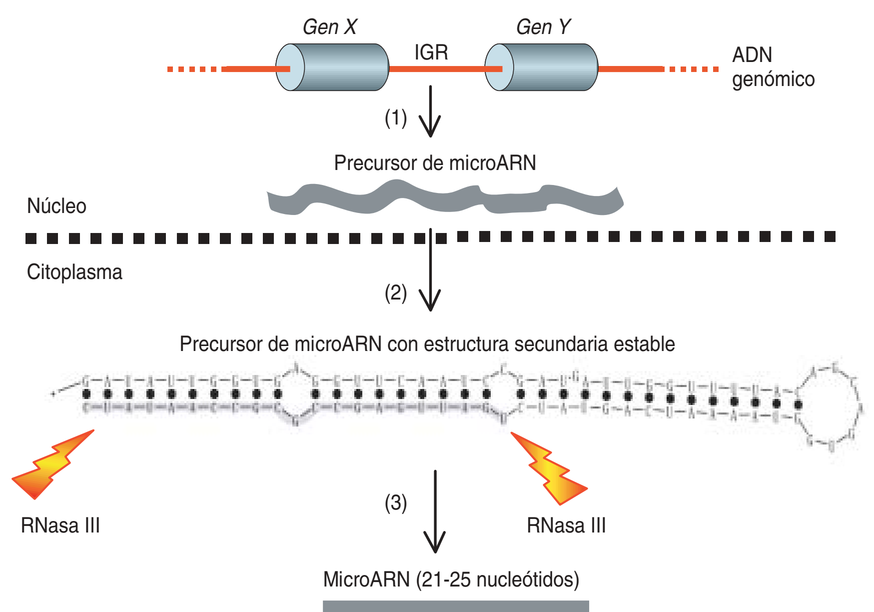
Otra posibilidad es que los microARN animales y vegetales interaccionen con los ARN mensajeros de los genes diana y afecten a otros procesos distintos de la traducción: la maduración del ARN mensajero o la localización y estabilidad del ARN. En este sentido, algunos microARN animales se emparejan con dominios, de aproximadamente 8 nucleótidos, situados en regiones 3' no codificadoras de los ARN mensajeros. Dichos dominios están implicados en la represión post-transcripcional de la expresión de genes diana; se encargan, sobre todo, de la estabilidad del mensajero y de la eficacia traduccional.

Recientemente, el grupo de James C. Carrington, de la Universidad estatal de Oregón, demostró que algunos microARN de plantas están capacitados para dirigir el procesamiento específico de ARN mensajeros diana y hacerlo de forma similar al modo en que los pequeños ARN interferentes fragmentan los ARN. Estos microARN se originan, a partir de regiones intergénicas del genoma de la planta, por medio de transcritos precursores de ARN con estructura secundaria estable. La expresión de tales microARN se regula durante el desarrollo de *A. thaliana* y, también, de *N. benthamiana* y *O. sativa*; donde igualmente se generan a partir de regiones intergénicas dotadas de capacidad para formar transcritos con estructuras secundarias firmes.

Los microARN presentan homología de secuencia total o parcial con los ARN mensajeros codificadores de factores de transcripción asociados a procesos de desarrollo y crecimiento en plantas. Entre ellos se encuentran los codificados por los genes *SCARECROW*, relacionados con el control de la división radial en células de la raíz y la señalización hormonal y luminosa, por *CUP-SHAPED-COTYLEDONS2*, implicado en la formación del meristemo apical de la raíz, por genes de respuesta a auxinas y por genes vinculados al desarrollo floral, o por genes *PHABULOSA* y *PHAVOLUTA*, que regulan la iniciación del meristemo y el desarrollo foliar.

En este proceso, el ARN precursor o pre-microARN (con estructura secundaria) es reconocido por una RNasa III; la enzima lo corta específicamente para producir el microARN correspondiente. Este se integra en un complejo ribonucleoproteico (RISC) a modo de señal de especificidad, ya que determina el reconocimiento de los ARN mensajeros diana complementarios.

Fruto de la interacción entre el RISC y el ARN mensajero se produce el corte de la molécula de mensajero en la región de emparejamiento entre ambos. Aunque estos microARN parecen actuar de forma similar a como lo hacen los pequeños ARN interferentes, existe una diferencia notable en su biogénesis. Los pequeños ARN interferentes resultan del procesamiento de una molécula inductora de ARN bicatenario y, por tanto, se inscriben en una población numerosa de moléculas que se sintetizan a ins-



6. PRODUCCION DE microARN a partir de regiones intergénicas del genoma. La región intergénica (IGR), situada entre los genes X e Y, se transcribe independientemente y genera un transcrito precursor de microARN (*1*). Este se procesa para dar lugar a un precursor definitivo con estructura secundaria estable (*2*). La estructura bicatenaria del precursor es reconocida por una RNasa III, enzima que lo digiere y libera la secuencia del microARN (*3*).

CESAR LLAVE

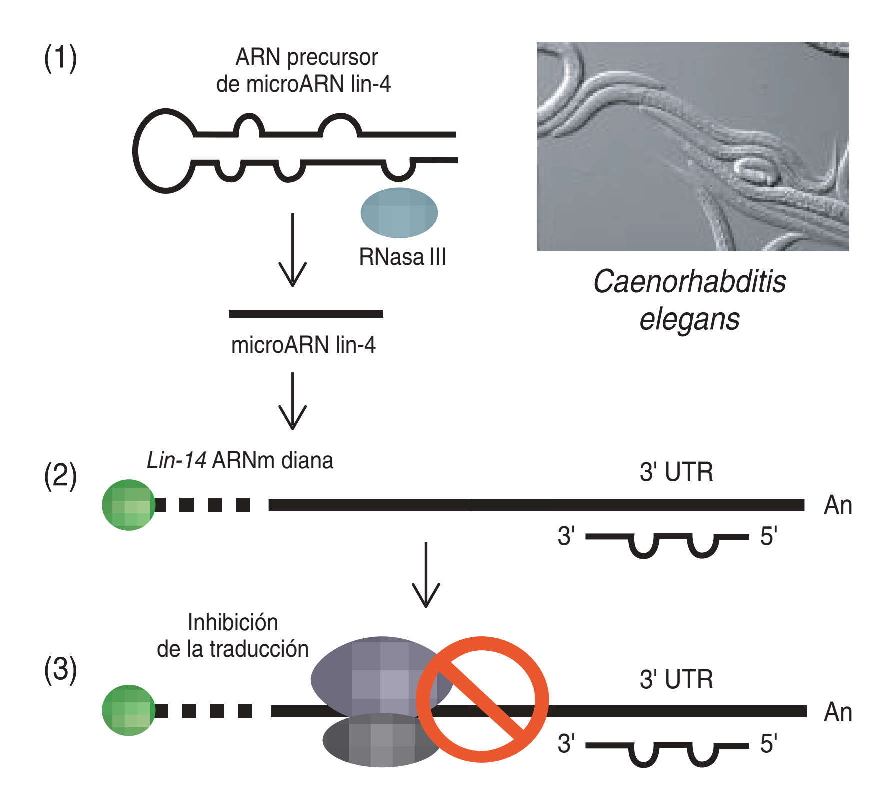
tancias de un fenómeno de silenciamiento. En otras palabras, cada molécula de ARN de doble cadena se procesa para dar lugar a multitud de pequeños ARN interferentes. Los microARN, por el contrario, poseen identidad propia y se originan independiente y específicamente a partir de una región intergénica del genoma de la planta; por eso, el procesamiento de cada ARN bicatenario precursor da lugar a un único tipo de microARN.

Incidencia en el desarrollo vegetal

Pero, ¿cuáles son las implicaciones del procesamiento, por microARN, de estos genes sobre el desarrollo de la planta? La mayoría de los genes diana cuya expresión se halla regulada por microARN se corresponde con factores de transcripción. Algunas de tales proteínas intervienen en el control del desarrollo de los meristemos. Este hecho sugiere que las rutas metabólicas dependientes de microARN pudieran estar implicadas en el control de la expresión génica durante el desarrollo de los organismos eucariotas.

Veamos un ejemplo ilustrativo de conexión entre ambos procesos: las mutaciones de *A. thaliana* en los genes *caf*, que codifica la RNasa III, y *hen 1*, de función desconocida. Las plantas mutantes presentan importantes defectos en el desarrollo; además, su maquinaria celular es incapaz de producir microARN. En consecuencia, todos los procesos reguladores que dependen de microARN se verán afectados en las plantas mutantes.

Los genes *PHABULOSA* y *PHAVOLUTA* ofrecen también un interés particular. Se caracterizan por presentar homología casi perfecta con microARN. Según parece, las plantas portadoras de mutaciones puntuales en la región del ARN mensajero de estos genes, que es complementaria con el microARN correspondiente, muestran considerables alteraciones fenotípicas. Estos ejemplos mencionados ponen de manifiesto que la actividad reguladora de los microARN es esencial para la correcta formación, diferenciación y desarrollo de los tejidos de la planta.



CESAR LLAVE

7. INHIBICION DE LA TRADUCCION de ARN mensajeros (ARNm) por pequeños ARN transitorios en el nemátodo *Caenorhabditis elegans*. El microARN *lin-4* (o pequeño ARN transitorio *lin-4*) se origina a partir de un precursor de ARN con estructura secundaria estable por mediación de una RNasa III (*1*). El microARN interacciona a través de un emparejamiento parcial de bases con la región 3′ no codificadora (UTR) del ARN mensajero diana (*2*); bloquea su traducción (*3*).

MicroARN y defensa antivírica

Los virus, parásitos obligados, necesitan utilizar los factores, procesos y estructuras del huésped para su propia multiplicación y consiguiente invasión sistémica. Estos factores, procesos y estructuras forman la base de las interacciones de compatibilidad entre el virus y el huésped que posibilitan el arraigo de la infección.

En el marco de interacciones virus-huésped, el segundo puede desplegar mecanismos moleculares que, como el silenciamiento génico, le permite bloquear algunas de las etapas del ciclo vírico y oponerse a la infección. Los virus de plantas, y al menos algún virus animal, han evolucionado bajo esta presión selectiva y han cifrado proteínas supresoras de la respuesta de silenciamiento génico, con la merma consiguiente de su eficacia protectora.

Se descubrió, hace poco, que algunas de estas proteínas supresoras de origen vírico se interponían en la actividad reguladora de los microARN. En determinadas plantas, las proteínas supresoras de silenciamiento inhiben el procesamiento, mediado por microARN, de algunos ARN mensajeros que codifican factores de transcripción. Tal labor obstructora se refleja en defectos del desarrollo de la planta: repercute en el crecimiento y la morfología. Con otras palabras, estas proteínas de origen vírico modulan la expresión de genes propios de la planta al entorpecer procesos básicos de su metabolismo celular.

Es posible que los virus, a través de sus proteínas supresoras, aprovechen esa obstrucción de la regulación génica mediada por microARN para activar o desactivar genes del huésped y así crear las condiciones favorables para la prosecución de su ciclo biológico.

Cabe también que las plantas sinteticen microARN de una manera espontánea, constitutivamente o en respuesta ante la infección vírica; en ambos casos, con la facultad de

reconocer secuencias víricas y promover su desactivación. De modo general, la expresión de proteínas supresoras durante la infección vírica inhibiría no sólo la regulación programada de genes de la planta, sino también cualquier alteración que los microARN ocasionaran en secuencias del virus.

Algunos de estos hipotéticos microARN complementarios con secuencias víricas podrían ejercer funciones similares a los microARN que regulan la expresión de los genes endógenos de la planta. No sólo instarían el procesamiento de los ARN víricos, sino que se dejarían sentir en su traducción, estabilidad y localización celular.

MicroARN y terapia génica

A diferencia de lo observado en las plantas, los microARN identificados en organismos animales no presentan un alto grado de homología con los correspondientes genes diana. Más bien se emparejan mediante complementariedad parcial de bases resultando en la supresión de la síntesis de proteína; sin embargo, los ensayos *in vitro* confirman que las células animales poseen la maquinaria bioquímica necesaria para realizar eficazmente el procesamiento de ARN mensajeros por microARN, siempre y cuando el grado de homología de secuencia entre ambos sea alto. La mayor parte de los genes regulados negativamente por microARN en los animales están implicados en procesos de diferenciación celular y apoptosis, organogénesis y desarrollo temporal.

Al grupo de Thomas Tuschl, del Instituto Max Planck de Química Biofísica, debemos la demostración de que la transfección de pequeños ARN interferentes en cultivos de células de humanos y otros mamíferos impedía la expresión de genes endógenos homólogos. El hallazgo afianzó el estudio funcional de genes en células de mamíferos y abrió las puertas a su eventual uso con fines terapéuticos.

Hoy, la expresión artificial de pequeños ARN en cultivos celulares se ha convertido en un método eficaz y específico para silenciar genes homólogos. Denominada, en su origen, silenciamiento génico mediado por ARN interferente, esta técnica innovadora se fundamenta en la desactivación dirigida de genes de interés por mediación de pequeños ARN sintetizados "in vitro".

8. MODELO DE PROCESAMIENTO DE ARN mensajeros mediado por microARN. El gen codificador del microARN alojado en la región intergénica (IGR/microARN) se transcribe y da origen a un precursor de microARN con estructura secundaria estable (*1*). Una RNasa III específica de ARN bicatenario reconoce y degrada al precursor (*2*). El microARN así generado se integra en un complejo enzimático (RISC), donde actúa como guía para la identificación de los ARN mensajeros (ARNm) diana (*3*). El complejo RISC-microARN provoca el procesamiento del ARNm para dar lugar a dos nuevas especies de ARN: 5′ ARNm y 3′ ARNm (*4*).

Los ARN así sintetizados operan de modo idéntico a como lo hacen los pequeños ARN interferentes o alguno de los microARN descritos en plantas: reconocen los ARN mensajeros mediante complementariedad de secuencia y guían su degradación.

Hasta la fecha, los pequeños ARN interferentes se han empleado con éxito para inhibir la replicación *in vitro* de diversos virus patógenos (poliovirus, virus VIH y virus de la hepatitis C). En numerosos estudios se ha comprobado que la transfección de estas pequeñas moléculas en cultivos celulares produce inmunidad frente a la infección vírica: inducen la destrucción de los ARN víricos. En el caso particular del virus del sida, la investigación *in vitro* reveló que el ARN vírico del complejo formado tras la entrada del virus en la célula (donde se forma un ADN vírico intermediario a partir del ARN vírico) se degrada rápidamente; la integración del ADN provírico en el genoma del huésped se bloquea en células inoculadas con pequeños ARN dirigidos contra diversas regiones del genoma vírico.

Además de utilizar directamente los virus como secuencias diana, otras estrategias alternativas emplean estos pequeños ARN para silenciar la expresión de otros ARN mensajeros celulares que cifran proteínas cruciales del ciclo vírico. Encontramos un ejemplo ilustrativo de ello en el control de la infección por VIH en células de mamíferos mediante el silenciamiento del ARN mensajero celular que codifica el factor CD4; esta proteína constituye el principal receptor de VIH en células humanas. La transfección de células T con pequeños ARN interferentes provocó la disminución de la expresión de CD4 en la superficie de la mayoría de las células; en consecuencia, cayó sustancialmente la tasa de infección por VIH. Se han cosechado resultados similares si tomaba por diana el receptor CCR5, un co-receptor necesario para la entrada de VIH en la célula.

Otro ejemplo prometedor del potencial de la aplicación de ARN interferente en terapéutica nos lo ofrece la protección contra la "hepatitis fulminante", enfermedad que se ma-

tores víricos capaces de expresar microARN de forma estable y en cualquier célula o tipo celular.

Habrá que superar otra barrera: el alto grado de especificidad en el reconocimiento de los mensajeros diana por parte de los microARN. Sépase que basta un solo cambio nucleotídico en la secuencia del gen diana para desarbolar por entero la respuesta de silenciamiento. Una situación cuya importancia se hace patente al observar la enorme diversidad de ciertos virus patógenos de humanos que, como el VIH, presentan altas tasas de error en su replicación y muestran una gran variabilidad genómica entre las personas infectadas.

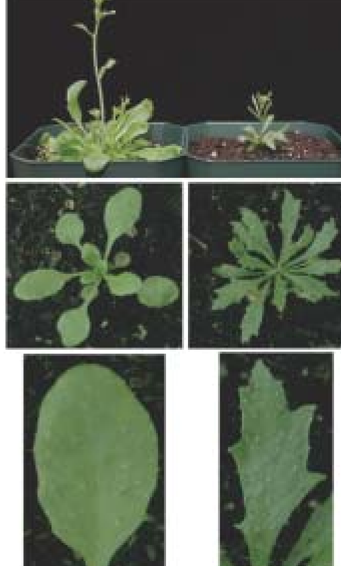
El estudio y conocimiento de estas pequeñas moléculas reguladoras nos han abierto las puertas a una nueva concepción en el entendimiento del control de la expresión génica.











9. SINTOMATOLOGIA INDUCIDA por un virus en plantas de *Arabidopsis thaliana*. Las proteínas supresoras de silenciamiento génico producidas durante la infección vírica se interponen en el metabolismo celular al alterar la expresión de genes de las plantas. Las plantas transgénicas que expresan constitutivamente proteínas supresoras muestran una sintomatología idéntica a la observada durante la infección por virus.

CESAR LLAVE; *FUENTE*: JAMES C. CARRINGTON, UNIVERSIDAD DEL ESTADO DE OREGON, EE.UU.

nifiesta en forma de fallo hepático severo. Los ensayos realizados en ratones han demostrado que el tratamiento con microARN contra receptor Fas situado en la superficie celular bloquea el desarrollo de la enfermedad y mejora la recuperación, aun cuando tales moléculas se apliquen después de la inducción de la enfermedad. El receptor Fas es un mediador de la muerte celular en el hígado, se produzca ésta por apoptosis o por necrosis. La reducción de la expresión de Fas en etapas precoces de la hepatitis podría prevenir su progresión hacia un estado más virulento de la enfermedad.

Una de las aplicaciones de la técnica de interferencia por ARN que mayores expectativas levanta es el empleo de pequeños ARN en terapia génica contra el cáncer en humanos. Algunos estudios se han centrado en la inhibición de la expresión de la telomerasa, enzima que se manifiesta en la mayoría de las células malignas pero no en las células normales. Estos trabajos confirman la capacidad de bloquear la producción de telomerasa en líneas celulares cancerígenas (carcinomas y sarcomas) en las que se habían introducido ARN interferente.

Sin embargo, la aplicación de esta nueva técnica de interferencia por ARN debe antes solucionar algunos problemas claves que limitan su uso con fines clínicos en humanos. El principal reto reside en la necesidad de desarrollar sistemas eficaces que permitan introducir directamente estos pequeños ARN en el tejido de interés. La mayoría de los métodos empleados resultan ineficientes o inadecuados. Algunos, como la transfección hidrodinámica, permiten aplicar estas moléculas en los órganos del ratón, si bien su puesta en práctica para uso clínico en humanos es inviable.

Otro problema que guarda estrecha relación con la aplicación de los pequeños ARN interferentes estriba en su estabilidad dentro del citoplasma celular. Estos microARN se hallan sujetos a degradación por enzimas celulares, por cuya razón el efecto silenciador sobre los ARN mensajeros desaparece al cabo de unas pocas generaciones. Para solventar el problema, algunos grupos están trabajando en el diseño de vec-

El autor

César Llave, adscrito al laboratorio de virología molecular de plantas del Centro de Investigaciones Biológicas del Consejo Superior de Investigaciones Científicas en Madrid, se doctoró en ciencias biológicas por la Universidad Complutense. Completó luego su formación en la Universidad estatal de Washington y en la estatal de Oregón. Su línea de investigación actual se centra en el estudio de las interacciones de compatibilidad y defensa entre plantas y virus, mediadas por microARN.

Bibliografía complementaria

GLIMPSES OF A TINY RNA WORLD. G. Ruvkun en *Science*, vol. 294, págs. 797-799; 2001.

AN RNA MICROCOSM. D. Baulcombe en *Science*, vol. 297, págs. 2002-2003; 2002.

RNA INTERFERENCE. G. Hannon en *Nature*, vol. 418, págs. 244-251; 2002.

REVEALING MICRO-RNAS IN PLANTS. L. Jones en *Trends in Plant Science*, vol. 7, págs. 473-475; 2002.

SIRNAS: A NEW WAVE OF RNA-BASED THERAPEUTICS. G. Coburn y B. Cullen en *Journal of Antimicrobial Chemotherapy*, DOI: 10.1093/jac/dkg166; 2003.

El torio, combustible nuclear

Un método poco explorado podría un día evitar que se fabricaran bombas con el combustible gastado

Mujid S. Kazimi

¿Cómo podría un grupo terrorista hacerse con el uranio o el plutonio necesarios para fabricar una bomba atómica? Esta pregunta ha pesado como una losa en la mente de muchos. Lo más fácil, probablemente, sería comprarlo; quizás a Corea del Norte que, según los servicios de inteligencia, podría contar con la infraestructura necesaria para producir una modesta reserva.

Hace un decenio, el mundo respiró tranquilo cuando los esfuerzos diplomáticos, incluidos los del ex presidente estadounidense Jimmy Carter, lograron desactivar lo que podía haberse convertido en un grave conflicto. Por entonces, Corea del Norte bloqueaba las tareas de control de la Agencia Internacional de Energía Atómica; incumplía así el compromiso que había asumido al firmar el tratado de no proliferación nuclear. En particular, los norcoreanos aseguraban haber producido sólo una pequeña cantidad de plutonio a partir del combustible gastado en sus reactores nucleares —insuficiente incluso para fabricar una sola bomba—, pero impidieron que los inspectores verificaran tal afirmación.

De entrada, ¿por qué Corea del Norte reprocesaba combustibles nucleares? La aplicación civil de la energía nuclear con fines pacíficos no requiere ningún procesamiento, según ha quedado patente en los países que la emplean. (En las centrales nucleares estadounidenses, el combustible gastado se acumula en recipientes secos o en piscinas de enfriamiento, en espera de su almacenamiento final en el depósito de la montaña de Yucca, en Nevada, que debería empezar a operar hacia el año 2010.) La decisión de Corea del Norte de reprocesar el combustible nuclear usado ¿no podía considerarse una prueba *prima facie* de que intentaba extraer el plutonio generado en sus reactores y usarlo para fabricar bombas nucleares?

No exactamente. Los reactores nucleares norcoreanos responden a un diseño bastante distinto del utilizado en los estadounidenses, que usan agua para las funciones de refrigerante y *moderador* (sustancia encargada de frenar los neutrones generados durante la fisión nuclear para que éstos inicien nuevas reacciones de fisión). Los reactores norcoreanos “magnox”, cuyo nombre proviene de la aleación de óxido de magnesio que rodea el combustible de uranio, se sirven de gas para refrigerar y de grafito para moderar. Con ligeras variaciones, se trata de un diseño que se viene aplicando desde hace tiempo en el Reino Unido.

El combustible gastado procedente de los reactores magnox no puede almacenarse directamente; antes debe reprocesarse para reducir su tendencia a oxidarse en presencia de aire o agua. Por tanto, el hecho de que Corea del Norte reprocesara su combustible gastado no podía, en sí mismo, considerarse una prueba de intenciones espurias. Sin embargo, su renuencia a permitir la entrada de inspectores internacionales constituía un motivo de preocupación.

El acuerdo promovido por Carter trajo cierto alivio. Para compensar el abandono de sus reactores de grafito, Pyongyang recibió de Washington asistencia técnica en la construcción de centrales nucleares del tipo de las que operan en EE.UU., junto con un generoso paquete de ayuda. La resolución tomada ofreció a los norcoreanos la posibilidad de desarrollar un programa de energía nuclear con fines pacíficos.

Por supuesto, sin una vigilancia adecuada nadie podía asegurar que los norcoreanos no estuvieran aprovechando sus nuevos reactores para generar plutonio, en algún lugar secreto, mediante el reprocesado de los combustibles gastados. Esa voluntad de trabajar clandestinamente en la obtención de materiales para la fabricación de armamento nuclear se hizo patente el año 2002, cuando se supo que Pakistán había enviado a Corea del Norte centrífugas de alta velocidad —equipos que se emplean en la obtención de uranio para bombas— a cambio de técnica de misiles. Obtener estos aparatos resulta cada vez más sencillo. En consecuencia, los intentos furtivos de transformar uranio ordinario en su forma enriquecida, adecuada para las bombas, resultan cada vez más difíciles de delatar. En

resumen, los programas secretos de armamento nuclear encierran una amenaza que, probablemente, siempre nos acompañará. Pero, ¿qué decir de la utilización de los combustibles gastados en la industria de guerra?

Una de las principales dificultades técnicas que entraña el reprocesado de los combustibles nucleares gastados proviene de su radiactividad. Estas sustancias siguen siendo intensamente radiactivas durante siglos después de la descarga; requieren, pues, manipularse en lugares remotos y en instalaciones equipadas con el blindaje adecuado para extraer el plutonio. ¿Existe alguna técnica que pudiera impedir el uso militar de estos combustibles gastados? Los expertos han venido reflexionando en torno a esta cuestión. Exploraremos aquí una posible respuesta que ha merecido creciente atención: el torio.

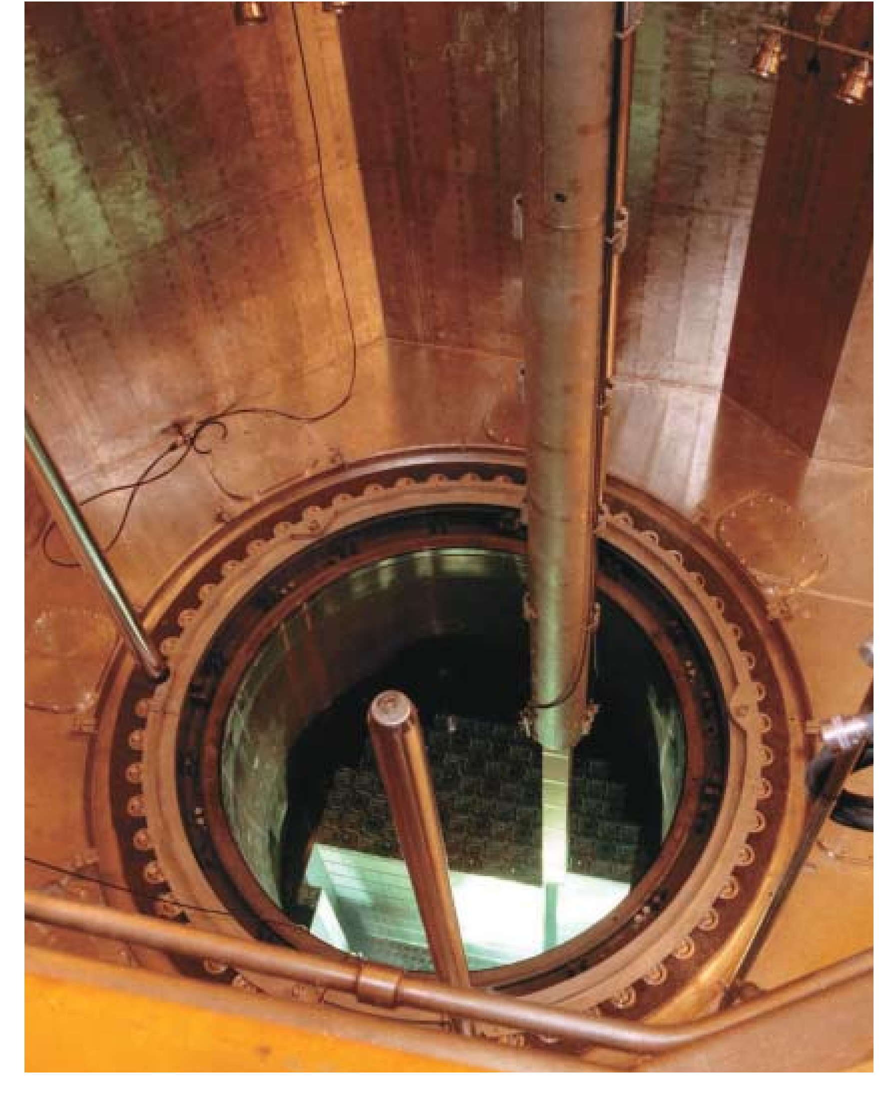
Las primeras aplicaciones nucleares del torio

Sobre el uso del torio en los reactores se ha especulado desde mediados del siglo pasado, cuando aparecieron las primeras centrales. No en vano se trata de un elemento que abunda en la corteza terrestre mucho más que el uranio. En números redondos, la cuantía de torio triplica la de uranio. Mas, para nuestro infortunio, resulta difícil inducir la división de los átomos de torio, condición indispensable para el proceso de fisión que tiene lugar en un reactor nuclear. Ahora bien, si se introduce en el reactor cierta cantidad de torio-232 (su isótopo natural), absorberá neutrones y se transformará en uranio-233 que, lo mismo que el uranio-235 (isótopo habitual usado para generar energía nuclear), alimenta las reacciones de fisión en cadena.

Por ello se dice que el torio, más que "fisible", es "fértil". En este aspecto se asemeja al uranio-238, que constituye más del 95 por ciento de la mayoría de los combustibles nucleares. Un reactor estándar produce distintos isótopos del plutonio a partir del uranio-238; parte de este plutonio experimenta, a su vez, un proceso de fisión, que se añade a la energía liberada por el uranio-235.

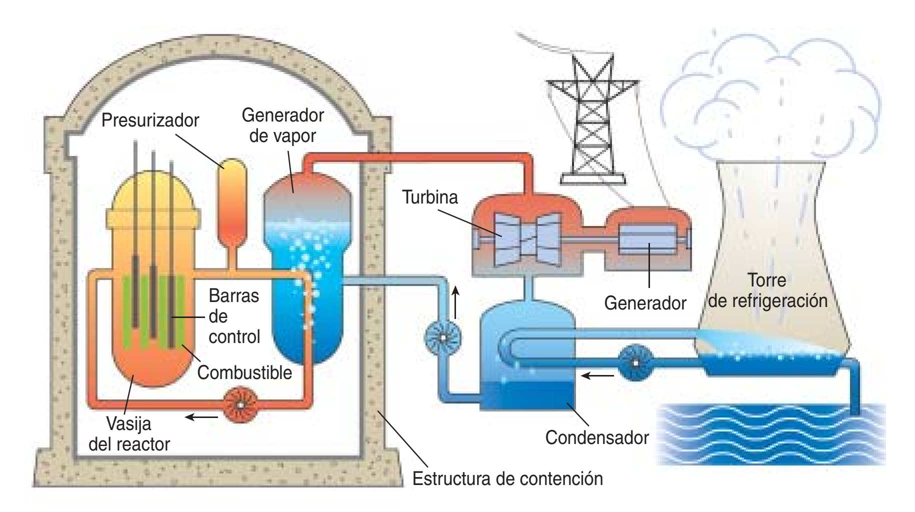
La aplicación del torio como combustible entraña algunas dificultades técnicas. La principal: éste no puede suministrar energía sin la ayuda previa de neutrones. Tales partículas podrían obtenerse de aceleradores, pero ello encarecería demasiado el proceso. Por el momento, la única solución práctica consiste en combinar el torio con combustibles nucleares usuales (formados por plutonio, por uranio enriquecido o por ambos) cuya fisión proporciona los neutrones que aquél necesita para iniciar su reacción en cadena.

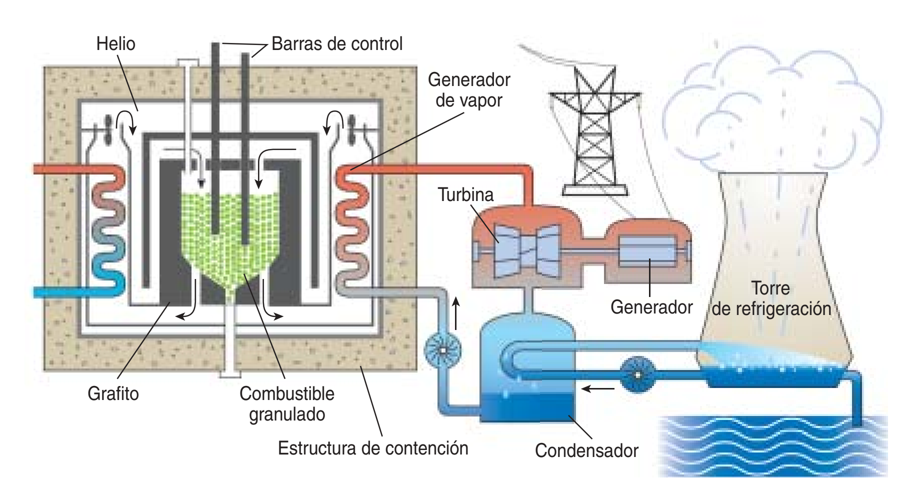
Pero trabajar con torio conlleva también ciertas ventajas. En primer lugar, la generación de uranio-233 a partir de torio resulta más eficiente que la de plutonio a partir de uranio-238. Ello se debe a que en el primer proceso se crean menos isótopos no fisibles. Merced a ese rendimiento, se limita la cantidad de combustible gastado por unidad de energía generada, con la merma consiguiente de residuos desechables. En segundo lugar, el dióxido de torio (la forma de torio que se utiliza en las centrales nucleares) constituye un compuesto muy estable; mucho más que el dióxido de uranio



1. DE URANIO SE ABASTECEN la mayoría de los 439 reactores nucleares del mundo que, en su conjunto, producen el 16 por ciento de la electricidad del planeta. En la mayoría de las instalaciones, las cápsulas de óxido de uranio llenan varillas que se disponen en una configuración cuadrada dentro de cada unidad de combustible. Un reactor moderno alberga unas 200 unidades de este tipo en su núcleo. Estas se substituyen periódicamente, para reprocesar o almacenar el combustible gastado. Uno de los inconvenientes de esta forma de generar electricidad es que sus residuos contienen plutonio, que puede extraerse (por medios químicos) del combustible usado y utilizarse para fabricar armas nucleares. Los combustibles que combinan torio y uranio reducen este peligro, dado que producen menos plutonio; además, la composición isotópica de éste no es la adecuada para fabricar bombas.

CORTESIA DE BRITISH ENERGY PLC





2. LOS COMBUSTIBLES BASADOS EN TORIO resultan apropiados para los reactores de agua a presión, que utilizan agua para transferir el calor del núcleo y frenar los neutrones generados en las reacciones de fisión (*arriba*), así como para los reactores de alta temperatura enfriados por gas, que utilizan un gas como el helio para transferir el calor y grafito sólido para frenar los neutrones (*abajo*). Multitud de reactores de agua a presión (incluido el primer reactor construido para la generación comercial de energía) operaron con torio durante varias pruebas iniciales. Asimismo, algunos reactores de alta temperatura enfriados por gas también se han ensayado con combustibles basados en torio; entre ellos, el alemán THTR-300, un reactor de 300 megawatts construido cerca de Hamburgo en los años ochenta. Este equipo respondía al diseño de "lecho fluido" indicado en la figura, en el que el combustible granulado se coloca en una vasija en forma de tolva. Esta disposición permite que la realimentación se realice de forma continua, evitando las costosas extracciones que periódicamente tienen lugar en la mayoría de las centrales nucleares, que deben pararse para recargar los reactores.

BARBARA AULICINO/*American Scientist*

empleado en los combustibles actuales. Por tanto, reduce el riesgo de que las pastillas de combustible reaccionen con el metal que las recubre, o con el agua refrigerante, en caso de producirse una brecha en la cubierta de protección. En tercer lugar, el dióxido de torio ofrece una conductividad térmica entre un 10 y un 15 por ciento mayor que la del dióxido de uranio, lo que facilita aún más la liberación de calor de las varillas o barras de combustible que se encuentran en el interior del reactor. En último término, el dióxido de torio se caracteriza por un punto de fusión unos 500 grados Celsius mayor que el del dióxido de uranio. Tal diferencia proporciona un margen de seguridad adicional en el caso de producirse un aumento transitorio de energía o una pérdida de refrigerante.

El conocimiento de estas ventajas ha estimulado la investigación sobre las aplicaciones nucleares del torio. Se han realizado incluso experimentos ocasionales en centrales nucleares en funcionamiento. Así, con el reactor Unidad Uno de Peach Bottom. Moderado con grafito y refrigerado con gas, este reactor instalado en el sudeste de Pennsylvania, operó, a mediados de los sesenta, con una combinación de torio y uranio enriquecido. Otro reactor refrigerado por gas ubicado en Fort St. Vrain, en Colorado, se alimentó con un combustible similar (basado en torio) entre 1976 y 1989. Durante los años sesenta se acometieron ensayos con mezclas sencillas de óxido de torio y óxido de uranio enriquecido, en reactores refrigerados por agua en los laboratorios de "BORAX", en Idaho, los de Elk River, en Minnesota, y en la central nuclear de Indian Point, estado de Nueva York. Entre 1977 y 1982 se utilizaron combinaciones más complejas de torio y uranio-235, o uranio-233, en un reactor refrigerado por agua de Shippingport, en Pennsylvania, en un programa experimental que pretendía desarrollar un combustible capaz de producir más material fisible del que consumía. La central nuclear de Shippingport, que empezó a funcionar en 1957, fue precisamente la primera que se construyó en EE.UU. para la generación comercial de electricidad.

Los estudios sobre las aplicaciones del torio en los combustibles nucleares no se ha restringido a los EE.UU. Los ingenieros alemanes, por ejemplo, han utilizado combinaciones de torio y uranio enriquecido, o de torio y plutonio, en reactores nucleares refrigerados con gas, así como con agua. Los combustibles basados en torio se han ensayado también en el Reino Unido, Francia, Japón, Rusia, Canadá y Brasil. Sin embargo, a pesar de su empeño inicial, la mayoría de estas naciones terminaron por abandonar, tiempo atrás, la vía del torio en sus centrales. El caso de la India, que sí ha persistido interesada, constituye una excepción. A mediado de los años noventa, este país em-

pezó a alimentar algunas de sus centrales nucleares con barras que contenían torio, para uniformizar la distribución de energía dentro del núcleo del reactor. Esta experiencia brindó a los ingenieros indios la oportunidad de comprobar la eficacia del torio en cuanto combustible. Los resultados positivos obtenidos les impulsó a decidirse por la aplicación de combustibles basados en torio en reactores más avanzados, todavía en construcción.

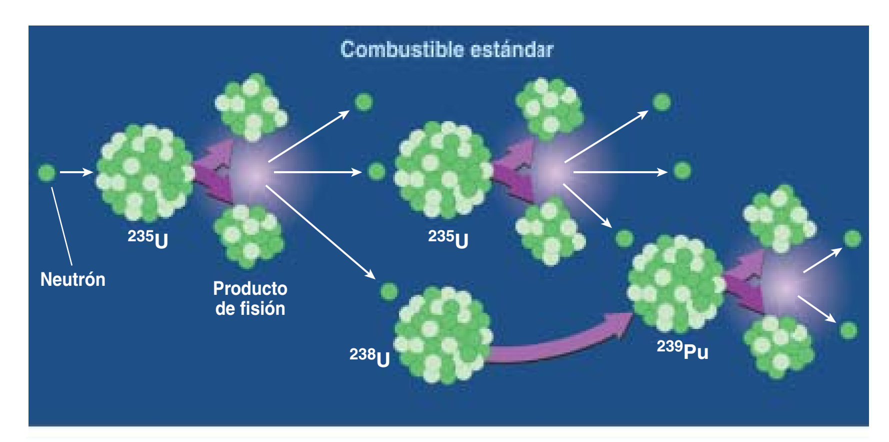
El interés de la India por los combustibles basados en torio arranca, en parte, de sus grandes reservas, estimadas en unas 290.000 toneladas, que le sitúan en segundo lugar después de Australia. Le permiten, pues, prescindir de las fuentes de uranio extranjeras. Pero la motivación principal no guarda relación con la balanza comercial, sino con la industria bélica. La India destina algunos reactores a la obtención de plutonio para bombas atómicas. De ahí que rechace las condiciones exigidas por los suministradores comerciales de uranio: Canadá, por ejemplo, obliga a sus clientes a dejarse controlar, para asegurar que el combustible (o el plutonio que se obtenga del mismo) no se derivará hacia armamento nuclear.

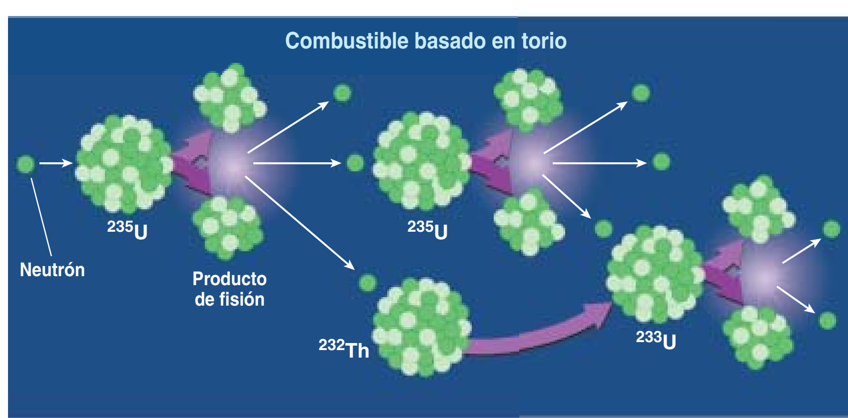
En otros países, los primeros ensayos con torio no condujeron a su adopción, por la razón poderosa de que su rendimiento en reactores de agua, tales como el primer núcleo de la central de Indian Point, no cumplía con las expectativas. A la vista de estos precedentes, puede resultar sorprendente que el torio vuelva a tomarse en consideración, aunque esta vez se conciba como un medio para frenar la potencial proliferación de armas nucleares. Para que el torio resulte eficaz en la prevención de la generación de plutonio, el combustible nuclear debe configurarse de forma distinta de la empleada en la mayoría de los experimentos del pasado. En esos ensayos se incorporaba uranio enriquecido (hoy descartado, ante la amenaza de la proliferación) y se daba por supuesto que el combustible gastado se reprocesaría para extraer su contenido fisible. Los combustibles basados en torio que se diseñan actualmente no contemplan ninguna de estas prácticas. Además, resisten mejor el calor y la radiación que se producen en el interior del núcleo de un reactor, lo que permite que una mayor parte del torio-232 fértil se convierta en uranio-233 fisible. En definitiva, pueden considerarse casi completamente distintos de los originales.

Cómo evitar la proliferación nuclear

Según he avanzado, la ausencia de uranio-233 en la naturaleza hace necesario el empleo de un material fisible distinto, como el uranio-235 (o quizá plutonio-239), para cebar el reactor que opere con torio. De acuerdo con la legislación vigente, que prohíbe el uso comercial de combustibles sumamente enriquecidos en uranio-235, el cebador debería incluir una cantidad considerable de uranio-238, no fisible; los estándares actuales requieren al menos un ochenta por ciento, y suele ser aún más. Por tanto, como ocurre en los reactores que operan en la actualidad, el combustible no puede usarse para fabricar una bomba si antes no se enriquece isotópicamente el uranio, proceso que entraña grandes dificultades técnicas.

La principal ventaja que ofrece la combinación de torio y uranio, frente a los combustibles usuales, radica en la notable reducción del contenido en plutonio del combustible gastado. La magnitud de tal reducción depende de la forma en que se combinan el uranio y el torio. Por ejemplo, si se mezclan homogéneamente en cada varilla de combustible, la cantidad de plutonio se reduce aproximadamente a la mitad. Sin embargo, ésta no es la única manera de combinar los dos elementos.





3. EL COMBUSTIBLE NUCLEAR ESTANDAR (*arriba*) contiene a la vez el isótopo fisible del uranio (^{235}U) y un isótopo no fisible (^{238}U). Cuando un neutrón incidente alcanza un núcleo de ^{235}U, éste se fisiona y libera dos o tres neutrones más. Estos, a su vez, provocan que otro núcleo de ^{235}U se rompa, o inducen a átomos de ^{238}U a transformarse en ^{239}Pu que, siendo él mismo fisible, contribuye también a la generación de energía. Los combustibles nucleares basados en torio (*abajo*) operan de forma bastante parecida; con una salvedad: en vez de producir plutonio a partir de ^{238}U, producen un isótopo fisible del uranio, el ^{233}U.

BARBARA AULICINO/*American Scientist*







4. PIONERO EN LA UTILIZACION de combustibles basados en torio, Alvin Radkowsky fue quizá su máximo impulsor hasta su muerte en 2002. Bajo la dirección del almirante Hyman Rickover, Radkowsky dirigió el Programa de Reactores de la Armada estadounidense. En 1954, lograron construir el primer submarino nuclear, el *USS Nautilus (abajo a la izquierda).* Aquí vemos a Radkowsky —con gafas— recibiendo un premio por logros técnicos de manos de Rickover (*arriba a la izquierda*). Detrás de ellos, un diagrama del primer reactor nuclear diseñado para la generación comercial de energía, que se construyó cerca de Shippingport, en el estado de Pennsylvania, a finales de los años cincuenta (*derecha*).

CORTESIA DE THORIUM POWER, INC. (*arriba, izquierda*); CORBIS (*derecha y abajo*)

En realidad, la mayoría de las investigaciones actuales optan por una combinación heterogénea; es decir, utilizan los dos elementos, pero los mantienen separados: el uranio cebador llena unas varillas de combustible y el torio otras. El principal impulsor de este concepto fue Alvin Radkowsky, quien, bajo la dirección del almirante Hyman Rickover, promovió la nuclearización de la Marina estadounidense durante los años cincuenta. Radkowsky realizó también notables aportaciones a la industria nuclear comercial durante los años sesenta y setenta. Por esas fechas, bajo el apremio de Edward Teller, que había sido profesor suyo y estaba entonces empeñado en encontrar una forma de evitar que las armas nucleares cayeran en manos inapropiadas, Radkowsky dirigió su atención hacia los combustibles basados en torio, que ya conocía por su capacidad de reducir los residuos nucleares. Radkowsky murió en 2002 a la edad de 86 años.

La idea de Radkowsky consistía en fabricar unidades especiales de combustible, que se pudieran emplear, con muy pocas modificaciones, en los reactores refrigerados por agua. Dichas unidades se dividen en una zona central, el cebador que contiene varillas de combustible llenas de uranio para reactores (es decir, con un máximo del 20 por ciento de uranio-235), y una zona de recubrimiento, con varillas de combustible llenas de torio y una pequeña cantidad de uranio. El hecho de tener uranio-238 en la cubierta evita que alguien retire las barritas y, sólo mediante tratamientos químicos, consiga separar el uranio-233 fisible que se crea con el tiempo (puesto que se encuentra mezclado con el isótopo ^{238}U).

Con el apoyo del Departamento de Energía de los EE.UU. (DOE) y la asistencia técnica del Laboratorio Nacional de Brookhaven, Thorium Power trabaja con el Instituto Kurchatov de Moscú para perfeccionar este método. Se pretende llenar las varillas de cebador con una aleación metálica y mantenerlas durante tres años en un reactor ruso, antes de reemplazarlas; las barritas de la cubierta se dejarán en el reactor durante 10 años. Sin embargo, sus resultados no serán directamente aplicables a las centrales nucleares de la mayor parte del resto del mundo, por los siguientes motivos: por un lado, el combustible no se halla en forma de óxido (lo usual en occidente) y, por otro, los reac-

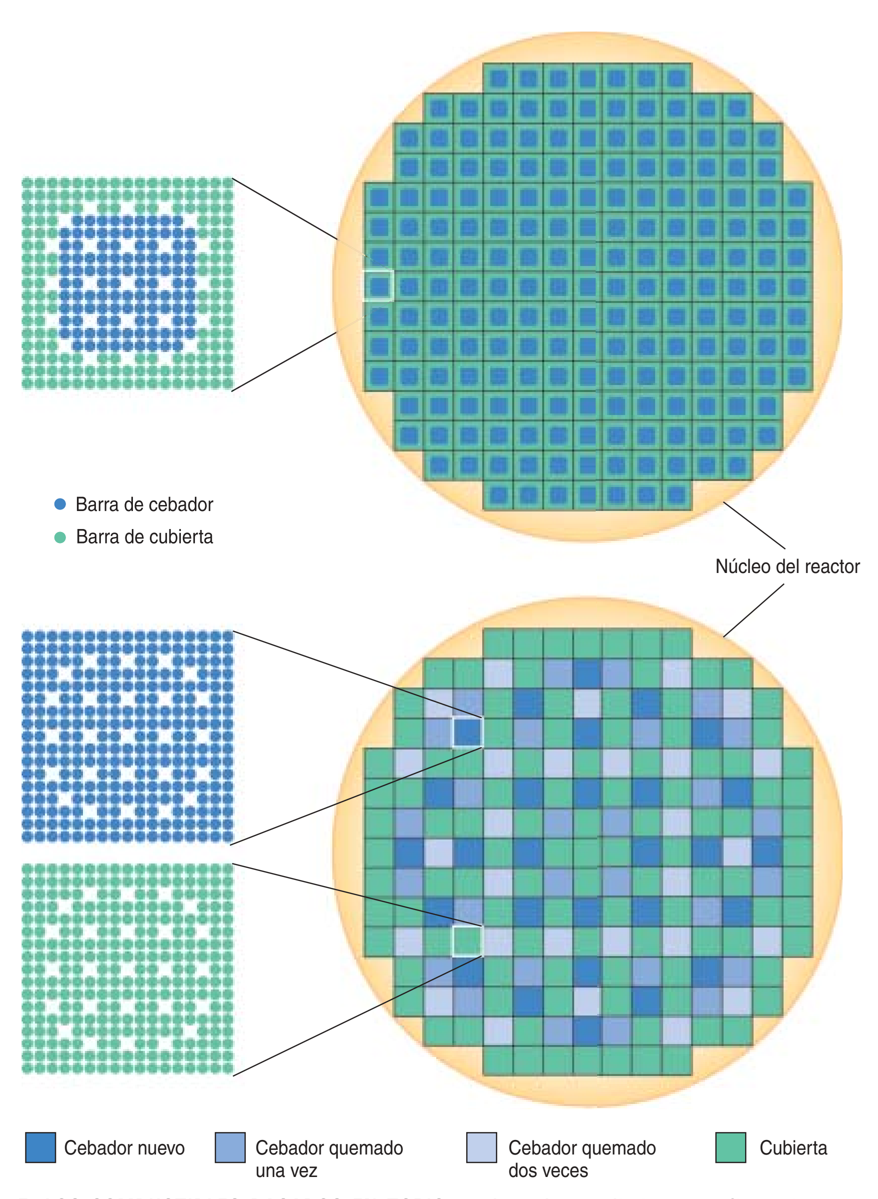
tores rusos que participan en estas pruebas usan una disposición hexagonal para las varillas de cada unidad de combustible, mientras que en occidente se encuentra más extendida la disposición cuadrada.

Radkowsky y sus colaboradores habían calculado que su esquema reduciría en un 80 por ciento la cantidad de plutonio producido, respecto a lo que ocurre en un reactor de combustible estándar que produzca la misma energía. Y lo que revestía mayor interés, encontraron que la mezcla de isótopos de plutonio generados, la mayor parte en el cebador, ofrecería pocas prestaciones armamentísticas, puesto que una bomba hecha con ellos ofrecería un escaso poder explosivo (probablemente se apagaría como "pólvora mojada"). Además, el plutonio presentaría un contenido tan elevado en su isótopo ^{238}Pu, que su calor de desintegración podría bastar para fundir o dañar los otros materiales utilizados en la fabricación del proyectil.

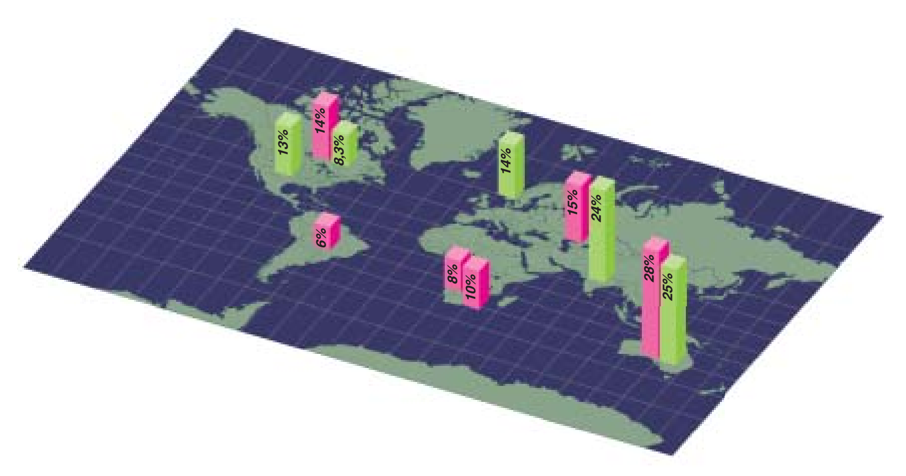
Incluso en el caso de que un grupo terrorista pretendiera usar el plutonio de la cubierta para construir una bomba aterradora (aunque no demasiado potente), extraerlo del combustible de torio de Radkowsky —en realidad, de cualquier combustible basado en torio— entrañaría más dificultades que extraerlo del combustible empleado actualmente. El combustible gastado de la cubierta contiene uranio-232, que en el transcurso de pocos meses se desintegra en isótopos emisores de rayos gamma de alta energía. Así, la extracción del plutonio requeriría protegerse mejor de la radiación y aumentar el uso de equipos accionados a distancia dentro de la planta de reprocesado, una complicación añadida a una tarea ya por sí difícil. Además, la abundancia de uranio-232 y sus productos altamente radiactivos en el combustible gastado frustrarían, a buen seguro, cualquier intento de separar del uranio-238 el uranio-233, que, al ser fisible, también podría servir para fabricar una bomba.

Los ensayos actuales

A la luz de las prestaciones que ofrecen los combustibles basados en torio, en cuanto a la disminución de residuos nucleares y prevención de la diseminación de materiales para fabricar bombas, no resulta sorprendente que el interés en tales combustibles haya vuelto a resurgir. El Departamento norteamericano de Energía se ha volcado en la promoción de su investigación. Además de financiar la empresa de Radkowsky y sus socios en sus pruebas con los reactores rusos, ha apoyado otros tres proyectos recientes. Entre ellos, el del consorcio que integran dos laboratorios nacionales (el Laboratorio de Ingeniería y Medio Ambiente de Idaho y el Laboratorio de Argonne), dos com-



5. LOS COMBUSTIBLES BASADOS EN TORIO pueden adoptar distintas configuraciones. En el diseño básico, concebido por Radkowsky, cada unidad de combustible (*cuadrados*) contiene dos tipos de varillas: en el centro las barritas ricas en uranio (o barras de cebador) y a su alrededor las barritas ricas en torio (o barras de cubierta) (*arriba*). El uranio, que incluye hasta un 20 por ciento del isótopo fisible ^{235}U, produce suficientes neutrones como para transformar el torio "fértil" que lo rodea en otro isótopo fisible del uranio, el ^{233}U. Esta combinación de combustibles en una misma unidad complica la realimentación del reactor, puesto que las barras de cebador deben reemplazarse con más frecuencia que las de cubierta. Un método alternativo, conocido como núcleo de "unidades completas" (*abajo*), utiliza unidades de combustible que contienen sólo barras de cebador (ricas en uranio) o sólo barras de cubierta (ricas en torio). Estas unidades se sustituyen con mayor facilidad.



6. LA INDIA CONSTITUYE EL UNICO PAIS que se dedica de forma activa a la obtención de combustibles basados en torio. En parte, ello se debe a que esta técnica le ofrece cierta independencia de los suministradores de uranio extranjeros. El país posee casi una cuarta parte de las reservas mundiales de torio; tiene, en comparación, poco uranio. En la gráfica se indican las naciones con, por lo menos, el 5 por ciento del uranio global (*rosa*) o del torio (*verde*). Las reservas de torio se conocen peor que las de uranio, lo que se explica por las escasas aplicaciones actuales de este elemento. (Cifras estimadas por la Asociación Nuclear Mundial.)

BARBARA AULICINO/*American Scientist*

pañías privadas fabricantes de combustible nuclear (Framatome Technologies y Westinghouse) y tres universidades (estatal Florida, Purdue y el Instituto de Tecnología de Massachusetts). El objetivo de la colaboración: dar con la forma de utilizar el torio en el reactor sin la complicación añadida de tener que operar con unidades combinadas de combustible (que mezclan barras de cebador y de cubierta), tal como ocurre en el diseño de Radkowsky.

El Laboratorio Nacional de Brookhaven y el Centro de Sistemas Avanzados de Energía Nuclear (CANES) del MIT participan en otro programa, cuya meta consiste en simplificar el diseño que separa el cebador y la cubierta en las unidades de combustible. Esto podría lograrse sustituyendo las "unidades combinadas" de Radkowsky (que contienen cebador y cubierta) por "unidades completas" (constituidas sólo por cebador o sólo por material de cubierta). Aunque mantenemos el término "cubierta", la metáfora resulta menos apropiada en este caso, puesto que las barras ricas en torio ya no "cubren" (o rodean) las barritas de cebador, sino que se alternan unas con otras en una disposición parecida a un tablero de ajedrez.

En un tercer proyecto de investigación, los ingenieros nucleares de Brookhaven y de la Universidad de Purdue examinaron el uso del torio cebado con plutonio en reactores de agua en ebullición. Estos reactores difieren de los que operan con agua a presión, más corrientes, que mantienen el agua de refrigeración a presión elevada, de forma que siempre permanezca líquida. Este programa pretende desarrollar una forma económica de consumir el plutonio sobrante de las bombas; y ello, sin producir otra generación de residuos de plutonio, tal como sucedería en el método de óxido mezclado, el más extendido actualmente. A este respecto, la investigación Brookhaven-Purdue guarda un parecido parcial con el trabajo que Thorium Power y sus socios rusos esperan acometer pronto.

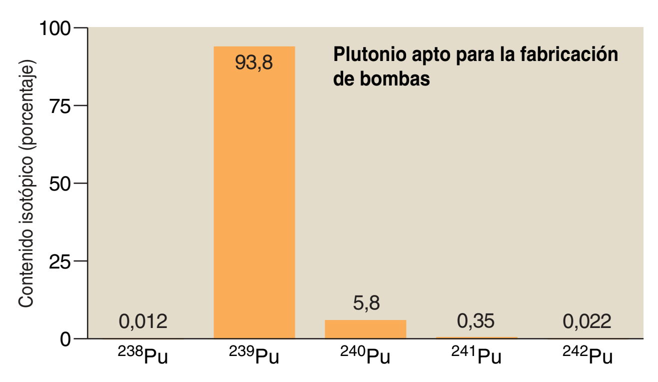
Durante los últimos años, en el CANES hemos dedicado un esfuerzo notable a evaluar las prestaciones de varios diseños, incluyendo distintas formas de combinar uranio y torio en una misma varilla de combustible. Me gustaría describir aquí algunos de nuestros resultados para las disposiciones con "unidades completas" (de tipo "tablero de ajedrez"), la estrategia que, desde mi punto de vista, ofrece mayores más posibilidades de éxito en el ámbito comercial.

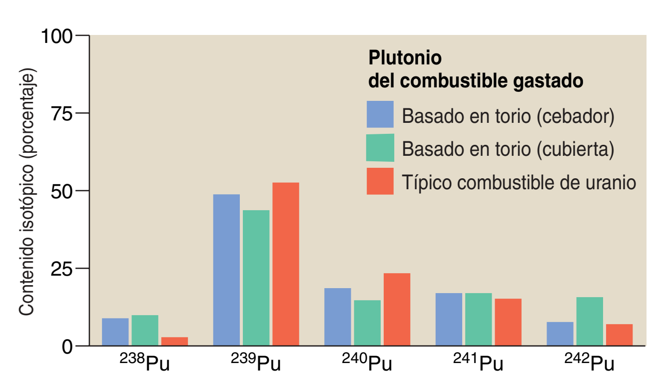
Balance final

Incluso con un núcleo de "unidades completas" (en el que cada unidad de combustible está constituida sólo por cebador o sólo por cubierta), la fabricación del combustible y su gestión dentro del reactor serían más complicadas que en un diseño estándar. En un reactor nuclear típico, las unidades de combustibles se mueven periódicamente, de suerte que cada una quede expuesta, en promedio, a las mismas condiciones de calor y radiación. En un núcleo de "unidades completas", en cambio, las unidades de cebador deben mantener niveles de energía significativamente superiores a la media, mientras que las de cubierta experimentan condiciones mucho menos exigentes. Por tanto, el combustible cebador alcanza temperaturas más elevadas, libera más productos de fisión gaseosos (al espacio que se les reserva dentro de las barras) y requiere más refrigeración que el combustible de cubierta.

Estas exigencias pueden atenderse de diversas formas. Por ejemplo, permitiendo que fluya más refrigerante a través de las unidades de cebador y preparando combustibles que opongan menos resistencia al flujo de calor. En la versión de Radkowsky-Kurchatov, las varillas de cebador constan de una aleación metálica de uranio (siguiendo diseños que han sido ensayados en submarinos rusos), lo que mejora su conductividad térmica. En el esquema propuesto por el equipo MIT-Brookhaven, las cápsulas de óxido de uranio del interior de las varillas de cebador están huecas, lo que rebaja su temperatura. Si bien las varillas de cubierta resultan menos problemáticas en este aspecto, exigen ciertas medidas: en concreto, un revestimiento exterior robusto, capaz de operar durante un tiempo de vida útil de hasta 13 o 14 años.

En el CANES también se han analizado las ventajas de los diseños de "unidades completas" en relación con su contribución en el esfuerzo por evitar la proliferación de materiales para fabricar bombas. Asimismo, se ha intentado evaluar su rendimiento económico. De acuerdo con nuestros resultados, estas disposiciones producen menos plutonio que las que mezclan el uranio y el torio a esca-





7. LOS REACTORES NUCLEARES generan plutonio como un subproducto, lo que aumenta las posibilidades de que este material pueda utilizarse para la fabricación de armas nucleares. Pero no cualquier plutonio sirve. El plutonio adecuado para las armas debe ser rico en el isótopo ^{239}Pu y ha de contener una pequeña proporción del isótopo altamente radiactivo ^{238}Pu (*izquierda*). Por tanto, el plutonio residuo de los combustibles basados en torio resulta mucho menos adecuado para la fabricación de bombas que el plutonio que se obtiene en un reactor estándar refrigerado por agua, puesto que contiene demasiado ^{238}Pu (*derecha*). Otra de las ventajas que ofrecen los reactores que operan con torio es que, globalmente, producen menos plutonio. Por ejemplo, un reactor alimentado con torio, con un núcleo de "unidades completas" (el tipo de reactor considerado en el gráfico de la derecha), produciría un total de 92 kilogramos de plutonio por gigawatt-año de electricidad generada, mientras que un reactor estándar refrigerado por agua produciría 232 kilogramos.

las más finas. Sin embargo, nuestros hallazgos no son tan optimistas como los primeros trabajos de Radkowsky: obtuvimos una reducción de sólo el 60 por ciento para el sistema de "unidades completas" y del 70 por ciento para las "unidades combinadas", mientras que Radkowsky estimó una reducción del 80 por ciento en la segunda opción.

Sin embargo, nuestros cálculos de la producción de plutonio confirman las predicciones de Radkowsky: el combustible gastado contendría una cantidad notable de plutonio-238, isótopo altamente radiactivo que generaría, por tanto, mucho calor. De hecho, el contenido en plutonio-238 triplicaría o cuadruplicaría el de combustibles de uranio estándares. Tal como señaló Radkowsky, el calor liberado por este isótopo haría bastante difícil, si no imposible, fabricar y mantener un arma nuclear.

La producción de tal cantidad de plutonio-238 se debe a que se consume (o se "quema", en el lenguaje de los ingenieros nucleares) una fracción del combustible mayor que en el caso de los reactores corrientes, que operan con uranio. Una cantidad equivalente de plutonio-238 podría generarse empleando un combustible de sólo uranio, pero ello requeriría una cantidad inicial de uranio fisible (^{235}U) muy superior a la habitual y, por tanto, un encarecimiento del proceso.

Así pues, nuestro estudio confirma que las dificultades técnicas pueden superarse y que los reactores que operan con torio podrían realmente frustrar los intentos clandestinos de utilizar el combustible gastado para fabricar bombas. Sin embargo, las conclusiones acerca de la viabilidad económica del torio son menos claras: estimamos que los combustibles basados en torio podrían costar entre un diez por ciento menos y un diez por ciento más que los combustibles nucleares usuales. Este amplio intervalo se debe a incertidumbres fundamentales acerca del coste del uranio cebador (que debe ser cuatro veces más enriquecido en uranio-235 que los combustibles nucleares normales), el coste asociado a la fabricación de las unidades de combustible y el ahorro derivado de la reducción de residuos.

Aunque parece poco probable que los criterios económicos conduzcan, por sí solos, a la adopción de los combustibles de torio, cabe considerar la ausencia de obstáculos técnicos de peso. Si bien es cierto que deberían modificarse las infraestructuras comerciales existentes, en principio, no se requieren nuevas técnicas. Y toda la experiencia acumulada en torno al uso combinado de torio y uranio enriquecido en reactores nucleares da verosimilitud a la posibilidad de extender las aplicaciones de ese método.



El autor

Mujid S. Kazimi se doctoró en ingeniería nuclear en 1973 en el Instituto de Tecnología de Massachusetts, donde ocupa la cátedra que patrocina la Compañía de Energía Eléctrica de Tokyo (TEPCO).

Bibliografía complementaria

THE NONPROLIFERATIVE LIGHT WATER REACTOR: A NEW APPROACH TO LIGHT WATER REACTOR CORE TECHNOLOGY. A. Radkowsky y A. Galperin en *Nuclear Technology*, vol. 124; págs. 215-222; 1998.

MEGAWATTS AND MEGATONS. R. L. Garwin y G. Charpak. University of Chicago Press; Chicago, 2002.

MICRO-HETEROGENEOUS THEORIA-URANIA FUELS FOR PRESSURIZED WATER REACTORS. S. Shwageraus, X. Zhao, M. Driscoll P. Hejzlar, M. S. Kazimi y J. S. Herring en *Nuclear Technology*, en prensa.

CURIOSIDADES DE LA FÍSICA

Jean-Michel Courty y Edouard Kierlik

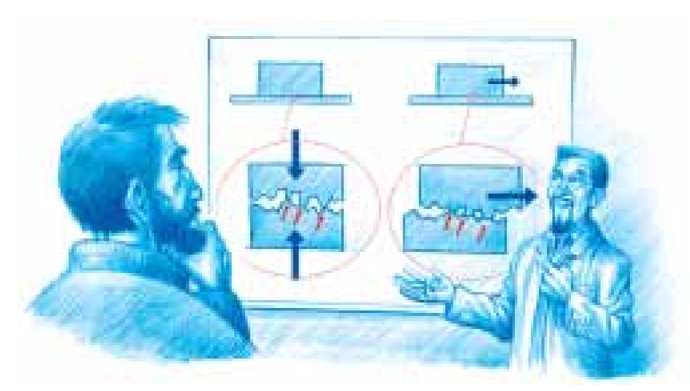
Sujeción perfecta

Aferrado al granito, un escalador es presa de la duda: el pie no le patina en el punto de la roca donde lo ha posado; pero, ¿puede apoyar en él todo su peso? Minutos después, su compañero de escalada emperna una plaquita de amarre a un pitón hincado en la pared y le invade una inquietud similar. Si resbalase, ¿aguantaría la plaquita? Las respuestas son sí y sí, tal y como garantizan las leyes del rozamiento seco. Analicemos de qué modo el rozamiento seco produce esas sujeciones perfectas.

Del empirismo a las leyes de Coulomb

Quien haya empujado una caja que repose en el suelo sabe que para conseguir moverla hay que ejercer un empuje que supere una cota mínima. ¿Cuál es la fuerza que contrarresta exactamente al empuje, mientras éste es pequeño? Se trata de la "fuerza de rozamiento seco", que se manifiesta cuando se intenta hacer deslizar entre sí las caras no lubricadas de dos sólidos. Contrariamente a los rozamientos fluidos, que sólo aparecen en el curso de un movimiento (resistencia del aire, etc.), el rozamiento seco se ejerce también sobre un cuerpo en reposo. Traspasado el umbral de rozamiento, el objeto se pone en movimiento y la fuerza de rozamiento actúa simplemente retardándolo.

No hay quien haga mudanzas que no conozca las extrañas propiedades de la fuerza de rozamiento seco: no importa que apile dos cajas iguales o las adose; para hacerlas deslizar deberá aplicar la misma fuerza. ¿Asombroso? Cuando las cajas están una sobre otra, el apilamiento presenta la mitad de contacto con el suelo que cuando están una tras otra. A principios del siglo XVI, el gran Leonardo da Vinci ya se interesó por el fenómeno. Sin embargo, fue Guillaume Amontons (1663-1705) el primer físico que formuló la ley del rozamiento seco: la fuerza necesaria para poner en movimiento relativo dos sólidos en contacto mutuo es proporcional a la fuerza que comprime uno contra otro e independiente del área de la superficie de contacto.



1. A una escala lo bastante pequeña, toda superficie sólida es rugosa. En contacto, dos de ellas entrelazan sus rugosidades y se resistirán al resbalamiento en tanto que éste no las deforme lo suficiente.

En 1785, Charles Coulomb (1736-1806) enuncia la ley que hoy lleva su nombre: el coeficiente de proporcionalidad entre la fuerza de rozamiento y la fuerza de presión depende del material. Ese coeficiente nunca rebasa la unidad, pasando desde 0,4, para papel con papel, a 0,5 para madera con madera, y llegando a 1 para una suela de caucho sobre piedra.

Durante mucho tiempo, la independencia de la fuerza umbral de rozamiento con respecto al área de la superficie de contacto fue un misterio. La causa no fue descubierta hasta el decenio de 1940 por los físicos británicos Frank Philip Bowden y David Tabor. Estudiando el contacto sólido a escala mesoscópica (intermedia entre la microscópica y la macroscópica) repararon en que dos sólidos que se tocan... casi no se tocan. Incluso la superficie mejor pulida presenta unas asperezas minúsculas cuyo tamaño típico es de una décima de micra. Esas rugosidades son las primeras zonas de contacto cuando se juntan ambos sólidos, y así se quedan. Por ello, el área de contacto real entre los dos sólidos es muy pequeña, como la de un faquir en su lecho de púas o la de un cepillo descansando sobre sus cerdas.

Contacto aparente y contacto real

La presión sobre las asperezas es igual a la fuerza compresora dividida por el área de contacto real. Esa presión es tan fuerte que supera largamente la resistencia de cualquier material: las asperezas se aplastan, aumenta el área de contacto y la presión disminuye hasta un valor límite, llamado dureza. En el equilibrio, la superficie real de contacto es igual al cociente entre la fuerza compresora y la dureza y por tanto no depende en absoluto de la superficie aparente de contacto. Así, para un cubo de cobre de 22 centímetros de arista y 90 kilogramos de masa que descanse en una superficie plana del mismo metal, la superficie de contacto entre el cubo y su apoyo es de un milímetro cuadrado aproximadamente. Esa superficie de contacto, es, por consiguiente, del orden de 500.000 veces menor que la superficie de contacto aparente.

Probemos ahora a hacer deslizar un sólido con respecto al otro. A la escala de las rugosidades, el deslizamiento es imposible, pues las asperezas se engranan. El movimiento no es posible mientras las asperezas no se hayan deformado hasta el punto de que el contacto se rompa. El deslizamiento de un sólido sobre otro es comparable al de un cepillo de hierro sobre una superficie metálica. Particularmente rígidas, las cerdas no resbalan, sino que se pliegan de momento antes de volver a desplegarse bruscamente para ir a bloquearse un poco más lejos. La fuerza necesaria para mover en vaivén uno de esos cepillos es proporcional al número de cerdas agarrotadas so-

BRUNO VACARO



2. Una bota tanto más resbala sobre un plano cuanto más inclinado es éste (ángulo *a*), pues la componente del peso paralela a la superficie, que la hace deslizar, crece con respecto a la componente perpendicular, que se aplica a la roca. La bota resbala cuando el cociente entre esas fuerzas excede del coeficiente de rozamiento.

3. Cuando los cuerpos están acelerados, las reacciones en negro no son opuestas a las fuerzas impulsoras de color naranja. Los rozamientos secos aseguran al alpinista incluso con un tornillo vertical del que se cuelgue con todo su peso; el martillazo asegura la imbricación de las superficies.

bre la superficie, número que aumenta cuando apretamos. Igualmente, la fuerza impulsora necesaria para desplazar un sólido en contacto con otro es proporcional al área representada por las asperezas engranadas, es decir, a la superficie real de contacto, y por ello, a la fuerza compresora. El examen microscópico resuelve la paradoja de la independencia aparente entre rozamiento y superficie de contacto: la superficie de contacto no es la que se creía. Ahora está resuelta la paradoja de las cajas: una superficie doble o una presión dos veces más intensa dan la misma superficie de contacto.

Sujeciones irreversibles

Ya podemos tranquilizar al escalador que pone la punta de su bota sobre un minúsculo plano inclinado de granito. ¿Puede apoyar todo su peso en ese punto? Sabremos si perderá pie dividiendo la fuerza que tiende a hacerlo deslizar (la componente de su peso paralela al plano inclinado) por la fuerza compresora (la componente del peso perpendicular al plano inclinado). El alpinista tantea ese cociente presionando levemente con el pie sobre su plano de apoyo. Si se adhiere, es que el cociente es menor que el coeficiente de rozamiento entre la superficie de su bota y la roca. El cociente en cuestión, igual a la tangente del ángulo *a* entre el plano inclinado y la horizontal, no depende pues en absoluto del peso del alpinista. Este, por tanto, puede confiar sin temor todo su peso al minúsculo plano inclinado y, con las manos así libres, afianzarse. El coeficiente de rozamiento del caucho sintético con la piedra seca es aproximadamente 1 y el alpinista no resbala mientras la inclinación del saliente de granito sobre el que apoya los dedos del pie no sea mayor que 45 grados. ¡Una pendiente considerable!

El segundo escalador puede también estar tranquilo. Si la tuerca que sujeta la plaquita de amarre no se desatornilla cuando él tire suavemente hacia abajo, tampoco se aflojará bajo una fuerza mayor. En el caso del tornillo, el equivalente a la pendiente es el ángulo que forma la rosca con el plano perpendicular al eje del perno. Al atornillar, todo ocurre como si se hiciera avanzar un objeto sobre un plano inclinado aplicándole una fuerza horizontal. La impulsión es fácil si la pendiente es lo bastante reducida para que la fuerza propulsora sea casi paralela al plano de deslizamiento; igualmente, el atornillado es fácil cuando el ángulo *a* es pequeño. Una vez atornillado del todo el tornillo o el perno, se le remata de un martillazo (o mejor un último cuarto de vuelta) al objeto de aumentar la imbricación de las rugosidades de su rosca con las del perno o la roca. Lograda esa inmovilización, toda tracción ejercida en la dirección del eje del tornillo o perno se traduce en una redistribución de fuerzas menores sobre cada uno de los elementos de la rosca. Mientras sea lo bastante pequeña la inclinación de ésta, las componentes de esas fuerzas paralelas al filete serán mínimas comparadas con las componentes perpendiculares al filete. Su cociente será, pues, inferior al coeficiente de rozamiento del metal con la roca (tornillo) o consigo mismo (perno). Así, existe un ángulo de rosca límite sin llegar al cual un perno no puede aflojarse, cualquiera que sea la tracción que sufra.

Ese "ángulo de deslizamiento" es del orden de 12 grados para los contactos de metal con metal y de unos 20 grados para los contactos de metal con madera, lo que explica por qué los pernos tienen una rosca menos inclinada que los tornillos para madera. La naturaleza irreversible (perfecta) de las sujeciones de tornillo se observa en los taburetes de los fotomatones. Un niño puede modificar su altura moviéndose de forma que roten en torno al husillo; un adulto de un quintal de peso se atrancará aunque alce los pies.

JUEGOS MATEMÁTICOS

Juan M.R. Parrondo

Matemáticas sostenibles

Alicia y Bruno quieren explotar un pequeño manantial de agua mineral. La productividad del manantial, es decir, el número de litros que se pueden extraer en un día, depende del uso que se le haya dado en los días anteriores. Si un día se extrae agua del manantial (supondremos que o bien no se extrae nada o se extrae la cantidad determinada por la productividad del día), la productividad disminuye en un litro diario. Si un día no se extrae nada y se permite así que la capa freática del manantial se recupere, su productividad aumentará en un litro diario, siempre que no exceda los 100. El primer día la productividad es la máxima, es decir, 100 litros diarios.

Los dos socios discuten sobre la forma óptima de explotar el manantial. Bruno opina que lo mejor es extraer agua siempre que la productividad no sea nula, pero Alicia pronto le demuestra que ésa no es una buena idea. Es cierto que extraerán agua los cien primeros días, pero en el día 101 la productividad será nula y no podrán extraer nada. Tendrían que esperar otros 100 días sin utilizar el manantial para que éste recuperase su productividad máxima de 100 litros diarios. Alicia se inclina por la siguiente estrategia: extraer agua un día sí y otro no. De este modo la productividad oscilará entre 100 y 99 y se extraerán 100 litros cada dos días, manteniendo una productividad media de 50 litros por día.

Sin embargo, después de hacer algunas cuentas Bruno volvió a defender su idea inicial. “Imagina —dice Bruno— que explotamos el manantial durante 100 días tal y como yo propongo, extrayendo agua todos los días. El primer día extraemos 100 litros y la productividad baja a 99 litros por día. El segundo día extraemos 99 litros y la productividad baja a 98 litros/día, y así sucesivamente. En el centésimo día, extraemos sólo 1 litro y la productividad baja a cero. En los cien días, habremos extraído un total de

$$100 + 99 + 98 + ... + 2 + 1 = \frac{(100 + 1) \times 100}{2} = 5050 \text{ litros.}$$

Ante la expresión de extrañeza de Alicia, Bruno le replica: “¿Ya no te acuerdas cómo se suma una progresión geométrica, como 1, 2, 3,..., 98, 99, 100? La suma es igual al número de sumandos multiplicado por la media del primero y el último.”

“Calculemos ahora —prosigue Bruno— la producción total utilizando tu política de extracción cada dos días. En este caso extraemos 100 litros el primer día, 100 litros el tercero, 100 el quinto, etc. Al final, habremos utilizado el manantial 50 días. Por tanto, la producción total será de 50 × 100 = 5000 litros, que es menor de lo conseguido con mi estrategia.”

“Sí —responde Alicia—, pero a partir del centésimo día tienes el manantial seco como la mojama. Te puedo demostrar que la mejor política *sostenible* es la mía. Por sostenible entiendo que el manantial se va a utilizar por un tiempo indefinido. Imagina que extraemos agua n días seguidos, dejamos descansar el manantial otros m días seguidos, y repetimos esta secuencia continuamente. Está claro que n tiene que ser igual a m. Si fuera mayor, la productividad acabaría siendo nula y, si fuera menor, estaríamos dejando de extraer agua incluso con productividad máxima. Es decir, tenemos que extraer agua n días y dejar descansar el manantial otros n días para que recupere su productividad inicial de 100 litros diarios. Al cabo de esos $2n$ días, habremos extraído una cantidad de agua:

$$100 + 99 + 98 + ... + (100 - n + 1) = \frac{(201 - n)\,n}{2} \text{ litros.}$$

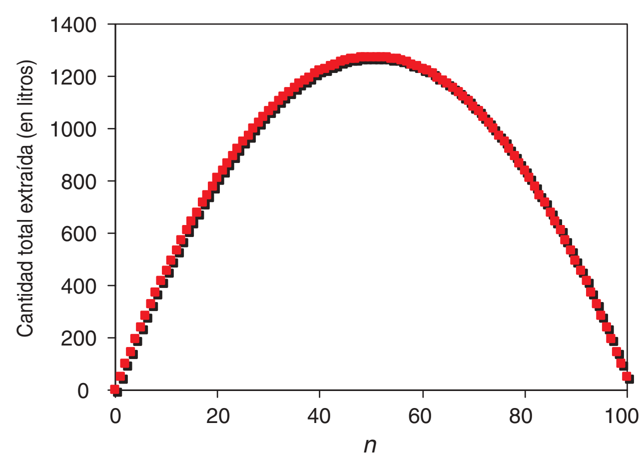
La productividad media se obtiene dividiendo esta cantidad total de agua por el número de días; es decir:

$$\frac{(201 - n)\,n}{2 \times (2n)} = \frac{201 - n}{4} \text{ litros/día;}$$

y es mayor cuanto menor sea n. El valor óptimo es por tanto $n = 1$, que coincide con la estrategia que yo defendí desde el comienzo: sacar agua un día sí y otro no y mantener la productividad siempre entre 100 y 99 litros por día”.

“Acepto la derrota —dijo Bruno. Aun así, mi estrategia es mejor que la tuya si tuviéramos que explotar el manantial sólo durante 100 días. ¿Cuál es entonces la mejor estrategia si el plazo de explotación es de N días?”

La pregunta de Bruno no es sencilla. Si el plazo es muy amplio, es decir, si N es muy grande, entonces parece claro que lo más conveniente es adoptar la estrategia de Alicia durante el comienzo de la explotación. Sin embargo, cuando el final del plazo se acerca conviene extraer más agua a pesar de disminuir la productividad del manantial, puesto que no nos preocupa cómo quede éste cuando el plazo de explotación venza. Analicemos entonces la siguiente estrategia: durante $N - n$ días utilizaremos la estrategia sostenible y en el resto, es decir, en los últimos n días, utilizaremos el manantial a diario. Llamaré *período sostenible* a los primeros $N - n$ días y *período de extracción exhaustiva* a los últimos n días. También tomaremos n de forma que $N - n$ sea par. De este modo, el último día del período sostenible no se extrae agua (recordemos que la estrategia sostenible es extraer el primer día, descansar el segundo, y así sucesivamente)



1. Cantidad total de agua extraída (obviando el término $50N$) en función del número de días n del plazo de extracción exhaustiva.

y el período de extracción exhaustiva comenzará con productividad 100 litros/día. El agua extraída en el período sostenible es

$$\frac{N-n}{2} \times 100 \text{ litros}$$

y la producción total en el período de extracción exhaustiva es

$$100 + 99 + 98 + ... + (100 - n + 1) = \frac{(201 - n)\, n}{2} \text{ litros.}$$

Por lo tanto, la cantidad total extraída con nuestra estrategia combinada es

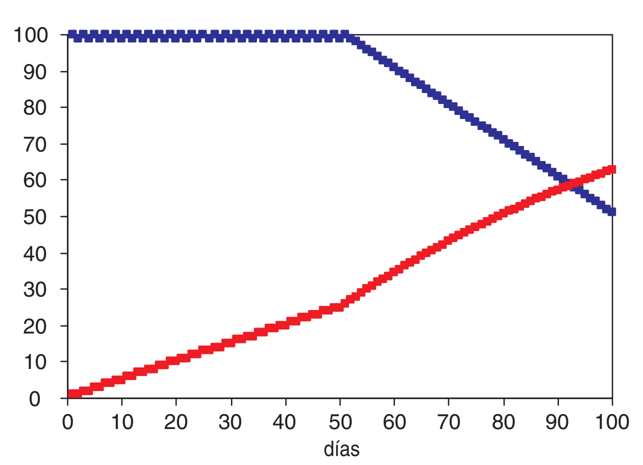
$$\frac{100\,N + 101n - n^2}{2} \text{ litros.}$$

Debemos encontrar el valor de n que hace máxima esta cantidad. En la figura 1 se muestra la cantidad total extraída (obviando el término proporcional a N), cuya gráfica es una parábola invertida. El vértice de la parábola está en $n = 50{,}5$ días. Encontramos por tanto que la cantidad total extraída alcanza su máximo tanto para $n = 50$ días como para $n = 51$ días. Para ambos casos, esta cantidad vale $50N + 1275$ litros. Por lo tanto, la mejor estrategia para un plazo de explotación de N días es extraer agua un día sí y otro no hasta que falten 50 (si N es par) o 51 días (si N es impar) para la finalización del plazo. A partir de ese momento lo que hay que hacer es extraer agua todos los días. Dejamos el manantial con una productividad de 50 litros por día, la mitad de su máximo valor posible. Pero esto nos importa poco: le tocará al siguiente esperar a que el manantial se recupere.

Para el caso de $N = 100$ días, el período sostenible durará 50 días y el de extracción exhaustiva otros 50. La cantidad total extraída es de 6275 litros, bastante superior a los 5000 de la estrategia sostenible y a los 5050 de la estrategia que proponía Bruno. En la figura 2 he representado la productividad y la producción total en función del tiempo. Vemos que la productividad oscila entre 100 y 99 durante el período sostenible y luego baja hasta 50 litros por día.

Pero, ¿es ésta la mejor estrategia posible? ¿No puede haber otras combinaciones con mayor producción? ¿Combinaciones en las que la productividad, por ejemplo, baje lentamente y no tenga el comportamiento abrupto de la figura 2? Yo no he sido capaz de encontrar ninguna, pero emplazo a los lectores a hacerlo o, en su defecto, a encontrar un argumento que pruebe que la estrategia de la figura 2 es la óptima.

Luis Dinís, de la Universidad Complutense de Madrid, Chris van den Broeck y Bart Cleuren, del Limburg Universitair Centrum (Bélgica), y yo mismo, hemos trabajado en estos problemas de optimización. Van den Broeck llama "estrategia avariciosa" a la de Bruno, puesto que trata de obtener el máximo posible cada día sin preocuparse de lo que ocurrirá al día siguiente. Dinís y yo también solemos referirnos a ella como "matar a la gallina de los huevos de oro". Ambas denominaciones indican que la estrategia de Bruno es una optimización de las ganancias inmediatas que resulta muy negativa a largo plazo, algo que desgraciadamente ocurre con frecuencia en la gestión política y empresarial de nuestro tiempo.



2. Productividad (*puntos azules*), en litros por día, y producción acumulada (*puntos rojos*), en miles de litros, en función del número de días transcurridos. En el día 50 se produce el cambio de la estrategia sostenible a la de extracción exhaustiva y la productividad comienza a bajar.

La combinación de una estrategia sostenible con un período de explotación exhaustiva puede darse en otras situaciones. Un atleta de medio fondo, por ejemplo, corre a una velocidad estacionaria la mayor parte de la carrera y realiza un *sprint* en los últimos metros, en donde realmente agota hasta la última de sus fuerzas. Un amigo ex atleta me informó de que un corredor emplea más de dos ritmos en una carrera debido a la presencia de los otros corredores. Sin embargo, en una contrarreloj, me aseguró, utilizaría sólo dos: el régimen "estacionario" y el *sprint*, como en la estrategia de la figura 2. Para el manantial, los últimos 50 días de explotación serían el equivalente al *sprint*: una explotación intensiva que acaba dejándolo medio exhausto.

parr@seneca.fis.ucm.es

Mark Fischetti

INYECCION DE COMBUSTIBLE

Combustión completa

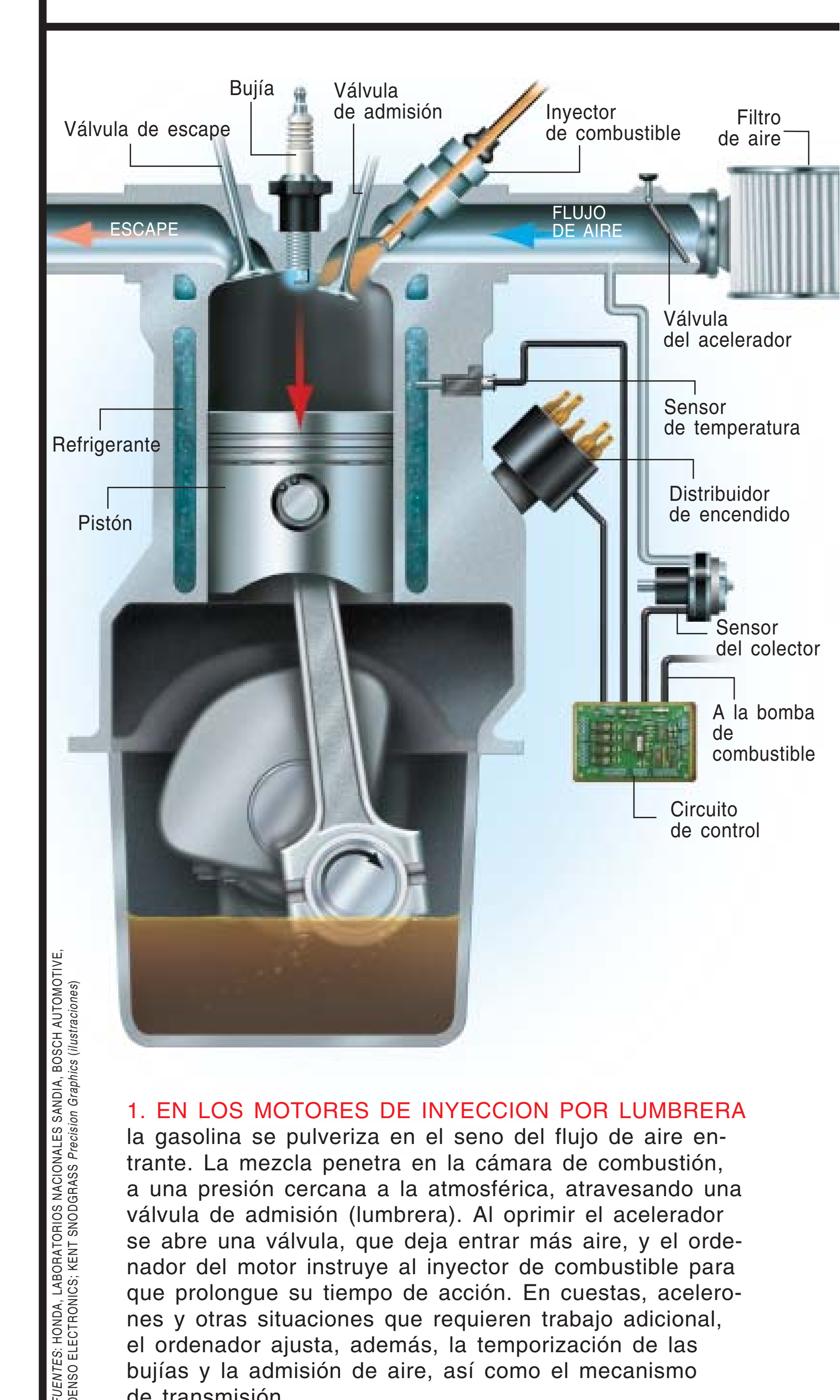
"¡Este deportivo es lo último! ¡Con inyección secuencial multipunto!". El vendedor trata de impresionarnos, pero lo cierto es que así operan prácticamente todos los turismos que se fabrican en la actualidad.

Durante decenios, el carburador actuó como un embudo que facilitaba la aspiración del aire y el combustible hacia los cilindros del motor. Las bujías encendían la mezcla y unas miniexplosiones impulsaban los pistones. El carburador funcionaba aceptablemente pero bregaba para controlar la relación de aire a combustible, e incluso para repartir equitativamente la gasolina entre los cilindros, con consecuencias adversas para el rendimiento del combustible, la generación de contaminantes y la suavidad del funcionamiento.

A finales de los setenta, la electrónica irrumpió en el sector automovilístico para regular la dosificación del combustible. Empezaron a instalarse unos inyectores en el colector de admisión, que venían a ser carburadores gobernados por ordenador. Pero el verdadero avance llegó a mediados del decenio siguiente, cuando se equipó a los nuevos modelos con inyectores multipunto —boquillas de inyección que distribuían con precisión el combustible en cada cilindro—. La gasolina se quemaba más uniformemente, con lo que aumentaban el rendimiento del combustible y la potencia útil, a la vez que disminuían las emisiones tóxicas. En la década de 1990, se implantó la inyección secuencial multipunto; el ordenador del motor controlaba cada inyector por separado, y no el conjunto de ellos, ofreciendo ventajas aún mayores. Cada uno de estos adelantos ha requerido mejoras en los inyectores, los sensores de realimentación y los procesadores.

La fuerza motriz que ha impulsado este largo progreso que partió del carburador reside en la mejora del control del combustible. La evolución se orienta ahora hacia la inyección directa, en la que la gasolina se envía directamente al cilindro sin antes mezclarse con aire. Los ingenieros están experimentando procedimientos que permitan mejorar la forma del pulverizador de combustible, así como arremolinar y estratificar la gasolina dentro del cilindro para que la combustión sea más completa y limpia.

Los motores de inyección directa funcionan a presiones mas elevadas, lo que algunos fabricantes consideran poco práctico, y producen óxidos nitrosos indeseables. Sin embargo, los primeros modelos ya están a la venta en Europa y Japón.



FUENTES: HONDA, LABORATORIOS NACIONALES SANDIA, BOSCH AUTOMOTIVE, DENSO ELECTRONICS; KENT SNODGRASS *Precision Graphics (ilustraciones)*

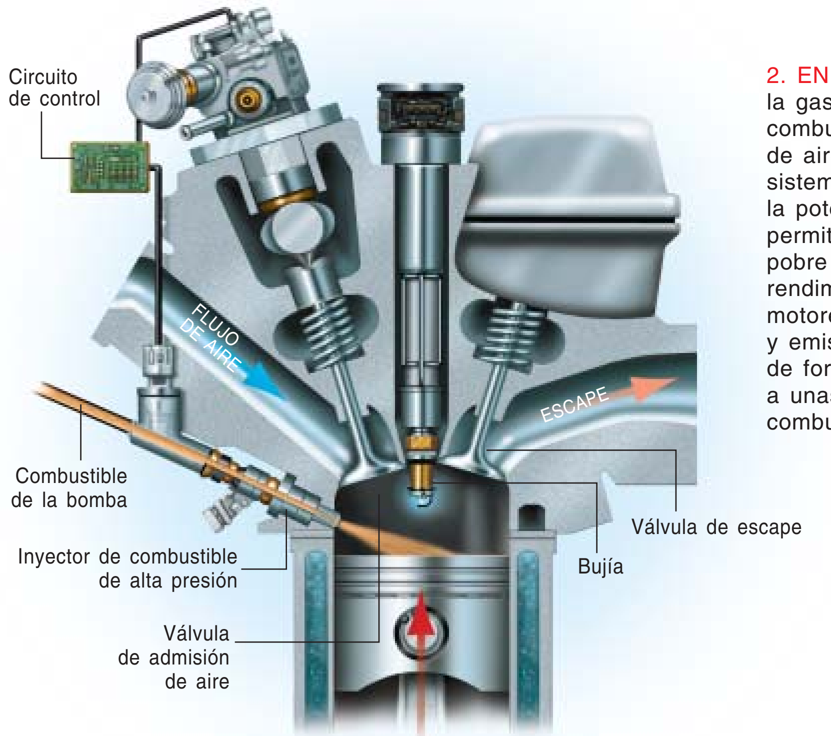
1. EN LOS MOTORES DE INYECCION POR LUMBRERA la gasolina se pulveriza en el seno del flujo de aire entrante. La mezcla penetra en la cámara de combustión, a una presión cercana a la atmosférica, atravesando una válvula de admisión (lumbrera). Al oprimir el acelerador se abre una válvula, que deja entrar más aire, y el ordenador del motor instruye al inyector de combustible para que prolongue su tiempo de acción. En cuestas, acelerones y otras situaciones que requieren trabajo adicional, el ordenador ajusta, además, la temporización de las bujías y la admisión de aire, así como el mecanismo de transmisión.

¿SABIA USTED QUE...?

➤ **CHISPAS:** Hay ingenieros trabajando en un prototipo de motor que opera mediante un "encendido por compresión de carga homogénea", que mejoraría el rendimiento del combustible y reduciría las emisiones de los diseños actuales. Se trata de una planta motriz que funciona con una premezcla de aire y combustible, como los motores de inyección de combustible, pero que se inflama debido a una presión elevada y no por la acción de una bujía, como los diésel.

➤ **POTENCIA DIESEL:** Los motores diésel recorren más distancia por litro de combustible que los de gasolina, porque el diésel contiene del orden de un 12 por ciento más de potencial energético que la gasolina. Además, funcionan con relaciones de compresión mayores (de 14:1 a 25:1, frente a de 8:1 a 12:1 para los de gasolina) con lo que extraen más energía de la combustión. Los motores diésel emplean la inyección directa para controlar con precisión el flujo de combustible y carecen de válvula de admisión, evitando las pérdidas que se producen en los motores de gasolina al forzar el aire a pasar por una válvula parcialmente cerrada; la potencia útil se controla modificando las dosis de combustible.

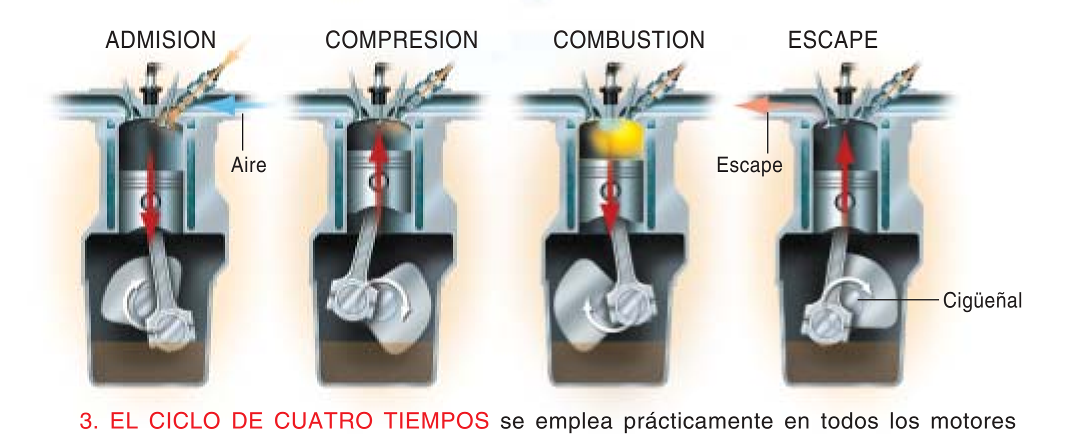
➤ **NEGOCIOS SUCIOS:** Los motores diésel generan más hollín que los de gasolina y requieren unos equipos de control de emisiones más complicados para cumplir con las normas estadounidenses. Los fabricantes se muestran reticentes a instalar esos procesadores debido a su coste añadido. Sin embargo, los motores diésel son muy comunes en Europa y Japón. Ello se debe, en parte, a una legislación más laxa en cuanto a niveles de contaminación y a un mayor precio de los combustibles; en estas condiciones, el mayor rendimiento en recorrido que ofrece el motor diésel compensa su coste.



2. EN LOS MOTORES DE INYECCION DIRECTA la gasolina se envía directamente a la cámara de combustión tras el cierre de la válvula de admisión de aire. La gasolina se vaporiza mejor que en los sistemas de inyección por lumbrera, aumentando la potencia útil; la temporización de la inyección permite al motor funcionar con una mezcla muy pobre de aire y combustible, aumentando así el rendimiento del combustible. Sin embargo, estos motores requieren técnicas avanzadas de control y emisiones. Los motores diésel también operan de forma "directa", pero el pistón comprime el aire a unas presiones lo bastante elevadas para que el combustible se inflame sin necesidad de una bujía.

PETER WITZE *Laboratorios Nacionales Sandia*

4. LA GASOLINA ATOMIZADA (*coloreada*) lanzada por un inyector de combustible incide en la válvula de admisión de aire cuando ésta se abre, en un experimento en los Laboratorios Nacionales Sandia.



3. EL CICLO DE CUATRO TIEMPOS se emplea prácticamente en todos los motores de automóvil para accionar un cigüeñal que transfiere potencia a las ruedas. La válvula de admisión se abre de forma que el aire y la gasolina entran cuando el pistón desciende. En su ascenso, el pistón comprime la mezcla de aire y combustible; una chispa la inflama. La fuerza que se genera en la combustión impulsa otra vez el pistón hacia abajo. La válvula de escape se abre para que el pistón, en su nuevo ascenso, expulse los subproductos.

Explicación científica

Modelos

CARL. G. HEMPEL. SELECTED PHILOSOPHICAL ESSAYS. Preparado por Richard Jeffrey. Cambridge University Press; Cambridge, 2000.

EXPLANATION. THEORETICAL APPROACHES AND APPLICATIONS. Dirigido por Giora Hon y Sam S. Rakover. Kluwer Academic Publishers; Dordrecht, 2001.

MAKING THINGS HAPPEN. A THEORY OF CAUSAL EXPLANATION, por James Woodward. Oxford University Press; Oxford, 2003.

EXPLANATIONS. STYLES OF EXPLANATION IN SCIENCE. Dirigido por John Cornwell. Oxford University Press; Oxford, 2004.

INFERENCE TO THE BEST EXPLANATION, por Peter Lipton. Routledge; Londres, 2004.

En el taller, el laboratorio, el observatorio o el campo, el científico atiende a hechos y procesos reales. Anota la trayectoria de este planeta o aquel cometa, describe los restos fósiles de este trilobite o aquella gimnosperma, identifica la función de esta enzima o de aquel neurotransmisor. Luego, en una segunda etapa, de reflexión, incluye lo observado o medido en una clase determinada de fenómenos o subsume el proceso bajo una ley o regularidad. Es decir, se propone ofrecer una explicación científica, dar el paso del conocer que existe *P* al comprender por qué existe *P*. Todos vemos que el cielo es azul, pero pocos comprenden por qué es azul. Al científico no sólo le interesa describir el mundo, sino también, y sobre todo, explicar por qué es tal cual es. Ese tránsito del *qué* al *por qué* constituye la entraña de su tarea.

Del concepto de explicación se han ocupado los filósofos desde la antigüedad clásica. Pero con la nueva epistemología nacida del entorno del Círculo de Viena y, con mayor exactitud, a raíz de su posterior refinamiento se ha convertido en uno de los temas centrales de la inquisición epistemológica. Cierto es que los neopositivistas rechazaban que la ciencia pudiera cuestionarse el *por qué* de las cosas, obsesionados por romper todo vínculo con lo que ellos reputaban metafísica idealista, causalidad incluida. En su opinión, la ciencia, un sistema de reglas para predecir los fenómenos, debía atenerse a la descripción de los mismos, no buscar su explicación. De ese corsé empezó a liberarse Rudolf Carnap, al admitir las causas físicas. Pero fue Hempel el que consolidó el giro neoempirista hacia esa nueva orientación (*Carl G. Hempel. Selected Philosophical Essays*). De lo que siguió podremos hacernos una idea con la breve muestra de libros que traemos a colación.

En toda explicación hemos de distinguir el *explanandum*, el fenómeno, proceso o regularidad a explicar, y el *explanans*, la causa, razón o ley explicativas. Si preguntamos: ¿por qué X?, aquí X constituye el *explanandum*. Si respondemos: "porque Y", aquí Y es el *explanans*. Para enhebrar uno y otro, en 1948, Hempel, en colaboración con Paul Oppenheim, publicó un artículo seminal, "Studies in the Logic of Explanation", en el volumen 15 de la revista *Philosophy of Science*. Hempel y Oppenheim ofrecían un primer esbozo de lo que se denomina modelo deductivo-nomológico (DN) de la explicación de hechos particulares. La idea general, sin embargo, la habían avanzado ya Aristóteles en sus *Analíticos posteriores*, John Stuart Mill, Karl Popper y H. Feigl. Pero ninguno la dotó del rigor ahora alcanzado. Hempel y Oppenheim se proponían aportar una exposición de la estructura lógica de "explica" que se equiparara en su alcance a la exposición de la estructura lógica de "es prueba de" desarrollada por Frege, Russell, Whitehead y Hilbert. La tesis recibió plena articulación en la monografía magistral de Hempel, publicado en 1965, *Aspects of Scientific Explanation and Other Essays in Philosophy of Science*. Sus postulados dominaron la epistemología de la segunda mitad del siglo veinte y, aunque han pasado a mejor vida, su apogeo y ocaso resultan instructivos.

Sostiene el modelo DN que toda explicación científica legítima implica la subsunción del hecho a explicar bajo una ley de la naturaleza. Toda explicación cabal, propone, tiene la estructura de un razonamiento deductivo, en cuyas premisas (el *explanans*) entran, con la ley o leyes, determinados enunciados verdaderos sobre las condiciones iniciales y datos de contorno, y la conclusión es el *explanandum*. Una buena explicación será, ante todo, un buen argumento.

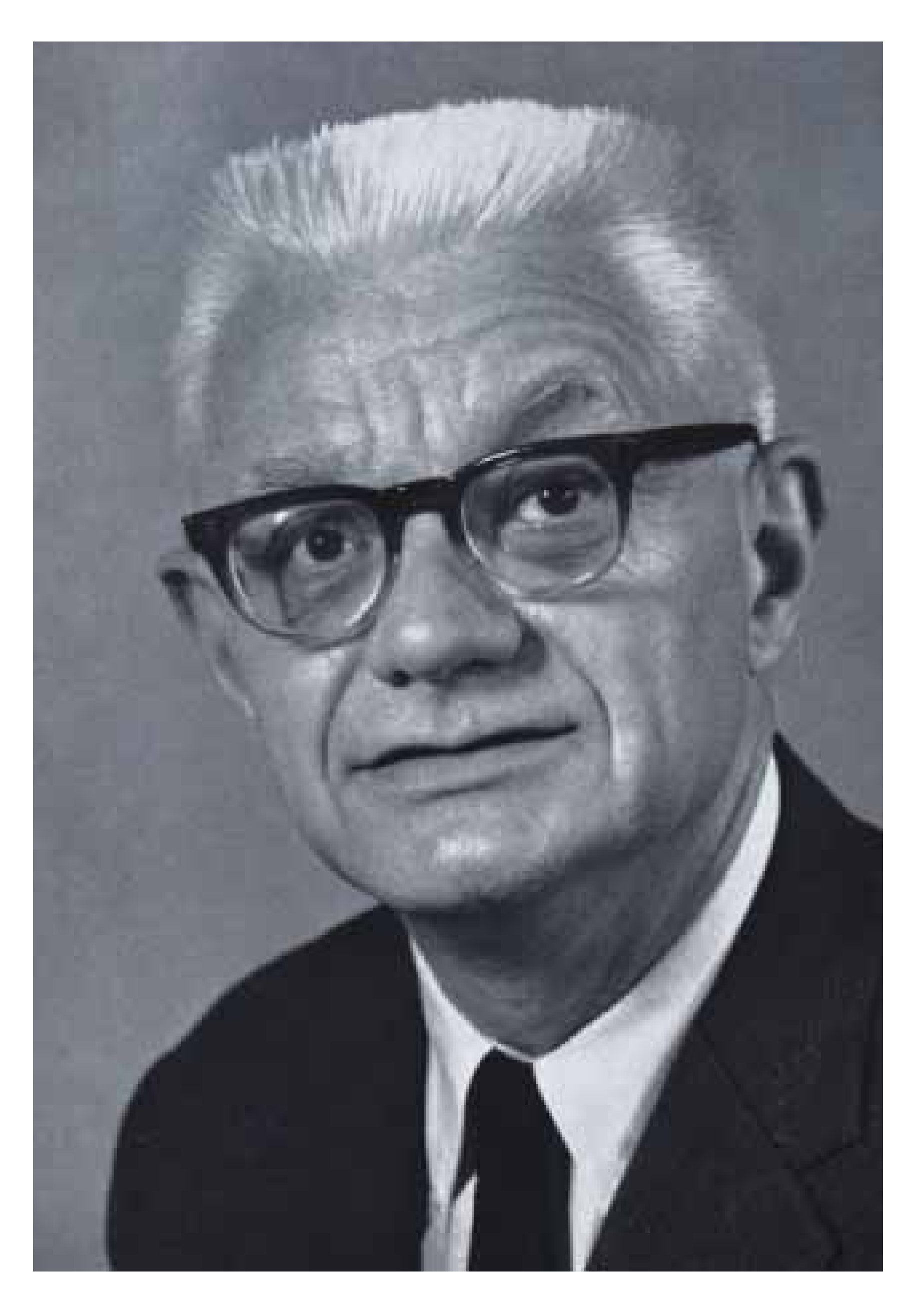
Puesto que nos encontramos con dos tipos de *explananda*, los hechos particulares y las regularidades generales, habrá también dos tipos de leyes que pueden aparecer en el *explanans*: leyes universales y leyes estadísticas. Para Hempel, sistematiza Wesley Salmon (*Explanation. Theoretical Approaches and Applications*), contamos con cuatro tipos de explicación: las explicaciones deductivo-nomológicas de hechos particulares por leyes universales; deductivo-nomológicas de regularidades generales por leyes universales; inductivo-estadísticas

de hechos particulares por leyes estadísticas, y deductivo-estadísticas de regularidades estadísticas por leyes estadísticas. Cuatro tipos de explicaciones que, en última instancia se resumen en dos: el modelo deductivo-nomológico (DN) y el inductivo-estadístico (IS).

En la estructura argumentativa, el modelo DN impone tres *condiciones lógicas* de adecuación que las explicaciones deben satisfacer. De entrada, su validez; esto es, el *explanans* ha de implicar deductivamente el *explanandum*. En segundo lugar, la deducción debe hacer un uso esencial de las leyes generales. Por último, el *explanans* ha de ser empíricamente contrastable. Súmase a estas tres condiciones lógicas una cuarta, *empírica*: las sentencias que constituyen el *explanans* han de ser verdaderas.

Muy pronto se demostró que la estructura lógica del modelo deductivo-nomológico carecía de carácter selectivo. Podía producir, por un lado, explicaciones privadas de sentido y, por otro, explicaciones de acontecimientos improbables. Con otras palabras, las tres condiciones lógicas de adecuación y el criterio empírico no son ni suficientes ni necesarios para una explicación convincente. En particular se contraargumentó que las explicaciones suelen presentar direccionalidad. Hay argumentos que sí pueden invertirse y seguir siendo buenos argumentos, pero las explicaciones, salvo en contados casos, no pueden invertirse. Por consiguiente, no todos los buenos argumentos que contienen leyes son buenas explicaciones. El ejemplo más socorrido es el de la sombra arrojada por el mástil de la bandera un día soleado. A partir de la información sobre la altura del mástil (h), el ángulo que el palo forma con el sol (ϕ) y las leyes que describen la propagación de la luz, se deduce la longitud (s) de la sombra. Esa inferencia constituye la explicación válida de s.

Mas, a partir de s, las mismas leyes y ϕ, se puede deducir h. Tal derivación, aunque aparentemente satisface todos los requisitos de una argumentación DN, no es una explicación de por qué el mástil tiene dicha altura. No lo es porque se olvida el papel desempeñado por la causalidad. La altura del mástil causa la longitud de su sombra, pero la longitud de la sombra es un efecto, no la causa de la altura del mástil; no se admite, en consecuencia, que sea buena explicación deducir h a partir de s. La actitud de Hempel ante la asimetría fue desconcertante. Para él, si la estructura argumentativa operaba en ambas direcciones entonces las dos direcciones eran legítimas, aun cuando la explicación procediera en sentido único.



1. Carl G. Hempel

El problema de la irrelevancia constituye otro punto débil del modelo DN. Se presenta cuando tenemos un argumento que satisface el modelo, pero parte del *explanans* no es un factor pertinente a la explicación. Suele mencionarse el ejemplo siguiente: "Todos los metales conducen electricidad. Todo lo que conduce electricidad está sujeto a la gravedad. Por tanto, todos los metales están sujetos a la gravedad." Aunque se trata de un argumento bien formado, con leyes en sus premisas, resulta manifiesto que la naturaleza conductora de los metales es irrelevante para su sujeción a la gravedad. Se pueden generar infinidad de ejemplos similares.

El modelo DN defiende, además, la "identidad estructural" de ex-

Cultura transversal

ELDERS. WISDOM FROM AUSTRALIA'S INDIGENOUS LEADERS, por Mandawuy Yunupingu y Lowitja O'Donoghue. Cambridge University Press; Cambridge, 2003.

En el seno de las llamadas poblaciones primitivas, los ancianos son tenidos como sabios y son reconocidos como receptores de las tradiciones sagradas, en especial las que guardan relación con los mitos de origen y con la interpretación de los tiempos naturales, o sea, de los que afectan desde la naturaleza la vida de quienes dependen de ésta para sobrevivir. Los ancianos están presentes en los ritos y suelen reunir el pensamiento colectivo en forma de memorias que transmiten de una generación a otra. De hecho, los ancianos ejercen el papel de bibliotecas vivas de los pueblos en los que han nacido.

Aunque este libro habla de relatos contados por ancianos, indudablemente no estamos refiriéndonos a un pueblo primitivo. En realidad, se trata de gentes viejas aculturadas y que, por lo tanto, ya no viven como lo hacían en tiempos preeuropeos. En el presente, aunque rechazan adoptar la identidad australiana, sin embargo, mucha de su cultura es una mezcla de esta última con la propiamente indígena. Y en muchos casos, una parte de sus individuos adultos vive en el interior de la sociedad australiana moderna, aunque no plenamente confundidos con los miembros de esta última. Ocurre, pues, que los rechazos sociales son mutuos, y a pesar de los múltiples mestizajes biológicos que se han dado entre ambas poblaciones, éstas procuran mantenerse separadas entre sí, por lo menos en el sentido de procurar confirmarse como étnica y nacionalmente distintas.

El texto que se ofrece en este libro está compuesto por nueve relatos orales, y han sido escritos a partir de conversaciones grabadas, luego transcritas en forma de pequeños capítulos que, unidos, forman una realidad de cultura transversal, esto es, común a todos los individuos que viven como indígenas en esta región del S. E. del continente australiano, ya en el estrecho de Torres, en islas correspondientes a Tasmania. Algunas de estas poblaciones fueron primero indígenas en el continente australiano, y después, debido a la presión europea, fueron siendo desplazadas hacia los puntos donde ahora se encuentran.

Los relatos fueron recogidos entre los años 2001 y 2002, y quienes proporcionaron la información eran líderes de clanes y de tribus, según los casos. En cada caso, las versiones que dieron de su cultura los informantes corresponden a los modos de vida que les eran tradicionales. Conciernen, por lo mismo, a temas y a creencias relacionados con las formas y prácticas que eran comunes en la región. En este contexto, se trata de modos de cultura cotidiana, y se refieren, por lo tanto, a tópicos que incluyen ideas sobre enfermedades y métodos de atender a su curación, a la filosofía que define los tratos con la salud y la atribución de lo bueno y lo malo a la voluntad de la Madre/tierra o de la Madre/naturaleza, fuente primera y última de la vida de los seres. El material filosófico que incluyen estos relatos, la reflexión que contienen, hay que entenderla en sus contrastes semánticos, o sea, en oposición a los propios de los símbolos que articulan el logos occidental.

El resguardo tópico asigna un espacio considerable a las técnicas de subsistencia tradicionales, al papel logístico de las respectivas ecologías del desierto y del mar. Así, los modos de explotar el desierto y de conseguir en éste recursos que los europeos no han sabido reconocer, el mar y sus secretos, la caza y la recolección, con sus respectivas estrategias rituales, el conocimiento del carácter cíclico de los bienes de la naturaleza, los espacios de actividad sexuada, la caza el hombre, la recolección la mujer. La vida de la familia, el desarrollo cohesivo de los vínculos derivados de la organización social del parentesco, y el sentido espiritual atribuido a la interpretación del papel de la naturaleza, son narraciones que impresionan por el cúmulo de simbolismos e iconografías que despliegan ante nosotros.

En el interior de los supuestos intelectuales de los narradores indígenas nos impresiona grandemente una de las reflexiones del más importante líder anciano a cargo de uno de los relatos. En especial, el comienzo de su experiencia con los europeos lleva consigo el traslado de la memoria antigua de su pueblo a nuestro presente, y es evidente que se trata de una historia traumática. Recordando ésta, nos dice que no puede haber reconciliación histórica entre los pueblos que los invadieron, los blancos, y los nativos que fueron destrozados por aquéllos. La diferencia está en el hecho de que los nativos caminaban por la tierra y la respetaban en sus productos, y así siempre fueron espirituales con la Madre naturaleza. En este respecto, la filosofía nativa no puede concebir la idea de que la tierra pueda comprarse y venderse y pretender, al mismo tiempo, ser sus propietarios.

Esta forma de insulto supone una culpabilidad por nuestra parte que sólo podemos redimir reconciliándonos con la Madre/tierra, pues tanto como ya no caminamos con ésta, dice el anciano, ha dejado de ser sagrada y, por lo tanto, ya no merecemos ocuparla por ser nosotros causa de ofensa material y espiritual permanente. Por esta razón, nunca podremos igualarnos a los blancos y es, por lo tanto, imposible que pensemos en la reconciliación con éstos mientras no conozcan y no respeten lo que nosotros amamos: la naturaleza con su correspondiente espiritualidad. De hecho, siempre estamos naciendo, y cuando quemamos ramas de un árbol, lo devolvemos a la vida por medio de la luz y el calor que nos proporciona su fuego. En la contraposición, y para el anciano, el cristiano es un hombre hecho para el hombre; para el anciano el hombre es un ser hecho para ejercer la espiritualidad con la naturaleza. Importan, por lo mismo, más las respuestas que las preguntas.

—CLAUDIO ESTEVA FABREGAT



2. ¿Por qué se caen las fichas? Pudiera haber diversas explicaciones de un fenómeno tan trivial.

plicaciones y predicciones. En ambos casos se articulan argumentos cuyas premisas establecen leyes de la naturaleza y condiciones iniciales. Con una sola diferencia: en la explicación conocemos ya que la conclusión del argumento es verdadera, mientras que esa eventualidad se desconoce todavía en la predicción. La estructura lógica que funde explicación y predicción no carece de aval histórico; constituye, de hecho, la estructura lógica del método newtoniano. Con la física de Newton se predijo la vuelta del cometa Halley en diciembre de 1758. Una vez se observó éste, el mismo argumento explica por qué retornó cuando lo hizo. Pero existen otros muchos casos en que la observación de un fenómeno nos permite predecir la observación de otro sin que el primero explique el segundo. La caída de la aguja de un barómetro nos permite predecir que habrá tormenta, pero no explica por qué. La teoría de la evolución aporta una explicación del registro fósil, pero no puede formular predicciones adecuadas sobre nuevas especies, porque el cambio evolutivo se halla sometido a variaciones estocásticas en las condiciones ambientales y en la morfología de los organismos.

Las explicaciones probabilistas ofrecen abundantes ejemplos donde la predicción y la explicación caminan por vías dispares. Cuando la probabilidad conferida por el *explanans* al *explanandum* es baja, no podemos predecir siquiera que el *explanandum* suceda en realidad, aunque a posteriori podamos explicar por qué aconteció. El modelo DN no sirve cuando recurrimos a las leyes estadísticas para explicar los sucesos. Para ese ámbito del conocimiento, Hempel propone el modelo de inducción estadística (IS); de acuerdo con el mismo, en los casos de explicación probabilística, las leyes más las condiciones auxiliares hacen muy plausible el *explanandum*. No le faltan tampoco contraejemplos a este modelo.

Con todas sus debilidades, nadie puede negarle atractivo a la idea de una estructura puramente lógica de la explicación. Y a su fuerza seductora hay que atribuir las dificultades que encontraron los diversos modelos sugeridos para paliar las deficiencias reseñadas. Las nuevas propuestas buscaban hacer más selectiva y pragmática la estructura lógica de la explicación científica. Una posibilidad era añadir una estructura objetiva. Esa es la postura que tomó Welesy Salmon con su teoría causal (*Explanation. Theoretical Approaches and Applications*) y ha extendido, recientemente, James Wooswerd (*Making Things Happen. A Theory of Causal Explanation*).

El modelo mecánico causal gira en torno a varios conceptos centrales. Ante todo, la noción de proceso causal. Por tal debe entenderse un proceso físico, como el movimiento de una partícula por el espacio, que se caracteriza por la capacidad de transmitir su propia estructura. Cuando un proceso causal interseca espacio-temporalmente con otro, modifican ambos su estructura. En la interacción causal producida en la colisión entre dos partículas, se altera la dirección y la energía cinética de una y otra. En la jerga, la explicación de un acontecimiento *A* reflejará los procesos y las interacciones causales que conducen hasta *A* —*aspecto etiológico de la explicación*— o al menos cierta porción de éstos, así como describirá los procesos y las interacciones que constituyen el propio proceso —*aspecto constitutivo de la explicación*—. La explicación revela, pues, que *A* se encuadra en un nexo causal.

Además, la dirección de la causación (la interacción entre la luz solar y el mástil causan la sombra) indica la dirección de la explicación. Explicar algo es describir qué es lo que lo causó. En general, la explicación científica aporta conocimiento de los mecanismos de *producción* y *propagación* de la estructura del mundo, pues debe reflejar el modo en que opera la naturaleza. A diferencia del modelo DN, el modelo del mecanismo causal no necesita apelar a leyes, ni tendría por qué estructurarse en argumentos. En su exposición más rigurosa, esta teoría deja, sin embargo, flancos desguarnecidos: si las causas no son reducibles a leyes y, por ende, a regularidades, ¿qué son entonces? Además, no faltan ejemplos en que una ley se utiliza para explicar otra sin mencionar la causalidad; así, la explicación de las le-

yes de Kepler mediante las leyes de la gravedad de Newton, que nada dicen sobre lo que causa los fenómenos gravitatorios. Algunos asimilan la explicación funcional, propia de la biología, a la explicación causal. En realidad defienden que las explicaciones adaptativas habituales en ese dominio no serían tales hasta que no descubra el mecanismo causal subyacente.

Frente a la teoría deductivo-nomológica, y en opción alternativa al modelo del mecanismo causal, Michael Friedman y Philip Kitcher idearon su propuesta, que bien pudiera considerarse complementaria del último: el modelo de unificación. Para ellos, las explicaciones científicas serían exclusivamente las que aporten una descripción del mundo en la que fenómenos dispares adquieren un sentido unitario y coherente. Las explicaciones son deducciones a partir de premisas que incluyen generalizaciones, sean o no leyes, y satisfacen una condición adicional *x* unificadora. Unificaciones de ese tenor han desempeñado un papel importante en la ciencia; sirva de botón de muestra paradigmático la unificación newtoniana de las teorías terrestres y celestes sobre el movimiento y la teoría unificadora de Maxwell de la electricidad y el magnetismo.

Para Thomas Kuhn, Peter Godfrey-Smith y otros la explicación opera de manera diferente según de qué disciplina estemos hablando e incluso de la época histórica dentro de una misma ciencia. Los criterios de una buena explicación en química no serían suficientes en física, ni la explicación de la afinidad en Lavoisier satisfaría la teoría del enlace hoy. Godfrey-Smith, en particular, es partidario de lo que él llama *contextualismo*. Tesis pragmatista en la que ha seguido su propio camino Bas van Fraassen, quien rechaza que la explicación sea algo inherente a la ciencia. Van Fraassen niega que el razonamiento científico incluya un juicio sobre el poder explicativo de las teorías. Para éste, la explicación constituye una relación triádica: el *explanans* y el *explanandum* se encuentran siempre en un *contexto definido*, un componente pragmático, que depende de las posibles contrastaciones que el investigador tiene en la mente.

En los últimos años se ha ido afianzando el modelo de inferencia en razón de la mejor explicación (IBE). Se trata de una regla de inferencia. En virtud de la misma, doquiera que dispongamos de un abanico de hipótesis posibles, todas ellas empíricamente adecuadas para dar cuenta de un fenómeno, hemos de admitir la verdad de la hipótesis que nos ofrece la mejor explicación del evento en cuestión, a tenor de los indicios, datos o pruebas. El modelo IBE hunde sus raíces en la teoría de la abducción avanzada por Charles Sanders Pierce (1839-1914). Esa teoría recibió una primera estructuración en 1965 cuando Gilbert Harman acuñó la expresión "inference to the best explanation" en un artículo con ese título aparecido en *Philosophical Review*.

Las inferencias lógicas acostumbran clasificarse en *inductivas, deductivas* y *abductivas*. En las deductivas, o demostrativas, las premisas

Filosofar

101 Philosophy Problems. Segunda edición, por Martin Cohen. Routledge; Londres, 2002.

Un problema, la existencia de problemas en filosofía, y que podía hacer el 102, se da por resuelto en esta obra que nace, precisamente, de esa solución. El autor justifica la existencia de problemas en filosofía, y nada menos que 101, cuando algunos autores clásicos, de admitir su existencia, no siempre clara y mucho menos de su posible solución, los limitaban a tres, a cinco... También considera que todas las cuestiones filosóficas de importancia se encuentran en el libro, aunque honestamente reconoce que hay otras que no lo son. En cualquier caso, el tratamiento de esas cuestiones se pretende lo más alejado posible de lo que considera "jerga académica" y en el libro se adopta un enfoque de la filosofía como actividad, como capacidad individual que hay que desarrollar y, para ello, hay que plantear problemas a resolver y no simplemente doctrinas más o menos sistematizadas que el lector llegue a asimilar pasivamente. Ello implica la exigencia al lector potencial de que no acepte sin más las cuestiones sino que se enfrente a ellas de manera crítica.

Con estas premisas el libro se divide en cuatro partes: Enunciado de problemas, Discusión o Comentarios a los mismos, Glosario alfabético con algunos términos y personajes citados y una Guía de lectura completada por un breve índice de nombres y conceptos. Los problemas filosóficos se plantean a través de cortas historias o relatos y van desde cuestiones lógicas paradójicas a las de filosofía natural pasando por juicios de valor, relatos morales, problemas filosóficos tradicionales que no le importan a nadie e incluso por imágenes también paradójicas que pertenecen más a los terrenos de la psicología de la percepción que a los de la filosofía. El libro constituye un buen intento de divulgar la filosofía en la línea que Martin Gardner hizo en matemáticas o Smullyan en lógica. La lectura y análisis de los problemas quizá tenga más atractivo que la de los comentarios. Tres reimpresiones de la primera edición en 1999 y dos de la segunda en 2002 corroboran el éxito de la obra que ha sido traducida al castellano en 2003. Y una cuestión final: ¿es verdad que todo lo que se dice en esta nota es falso?

—Javier de Lorenzo



3. "La explicación", composición de George B. Dieboold.

implican la conclusión. Aquí la conclusión no puede ser falsa si las premisas son verdaderas. Las inferencias *no demostrativas* pueden llegar a conclusiones falsas, aun cuando las premisas sean verdaderas. La inducción y la abducción constituyen formas de inferencia no demostrativas.

En su libro *Inference to the Best Explanation* publicado más de un cuarto de siglo después del artículo original de Harman, ahora actualizado en una segunda edición y esquematizada su doctrina en *Explanations. Styles of Explanation in Science*, Peter Lipton acometía una elaboración profunda del método. La ciencia, asevera, depende de los juicios sobre las pruebas en que se basa una teoría. Los científicos deben juzgar si una observación o el resultado de un experimento presta soporte, rechaza o es simplemente irrelevante para una hipótesis dada. De manera similar, los científicos podrían juzgar que, con las pruebas disponibles, una hipótesis tendría que aceptarse como correcta, rechazarse por falsa o ni lo uno ni lo otro. En ocasiones, estos juicios pueden hacerse por razones deductivas. Si un resultado experimental contradice una hipótesis, entonces la verdad de la prueba implica deductivamente la falsedad de la hipótesis de partida. Pero en la mayoría de los casos la conexión entre prueba e hipótesis no es demostrativa, sino inductiva. Ocurre siempre que se infiere que una hipótesis general es correcta en razón de las pruebas disponibles, si bien siempre cabe la posibilidad de que la hipótesis sea falsa, aun cuando los datos sean correctos. Apela al ejemplo de Darwin. Este infirió la hipótesis de la selección natural porque, aunque no estaba implicada por sus indicios biológicos, ese mecanismo evolutivo aportaba la mejor explicación de tales indicios. Cuando un astrónomo infiere que una galaxia se está alejando de la Tierra con una velocidad determinada, lo hace porque la recesión constituye la mejor explicación del corrimiento hacia el rojo del espectro de la galaxia observado.

El modelo IBE no ha quedado libre de críticas. Van Fraassen objeta que, puesto que existen muchas teorías ontológicamente incompatibles, aunque empíricamente equivalentes, resulta improbable que la teoría verdadera se encuentre en la clase elegida; por tanto, resulta improbable que la teoría mejor explicativa sea verdadera. Reta a los defensores del modelo a mostrar de qué modo podemos conocer que ninguna de las otras explicaciones que no hemos considerado es tan buena como la supuesta mejor. En opinión de van Fraassen, no sólo las falsas teorías aportan buenas explicaciones (la mecánica newtoniana es falsa, pese a dar una buena explicación de las mareas), sino que, además, la fuerza explicativa es, según hemos avanzado, una relación pragmática entre una teoría, un hecho y un contexto, donde el último está determinado por las ideas e intereses del investigador.

—LUIS ALONSO

AVENTURAS PROBLEMÁTICAS

Dennis E. Shasha

Salto-Raptor

Veamos un juego que utiliza un tablero de tres en raya y unas cuantas fichas que se colocan en ese tablero. Estas son sus sencillas reglas: se puede saltar sobre una ficha si ésta se encuentra entre otra y un cuadro vacío (a lo largo de cualquier línea vertical, horizontal o diagonal). Si un jugador hace saltar una de sus piezas sobre otra, ésta se retira del tablero, como en las damas.

Se puede jugar en solitario. En tal caso, el objetivo consiste en dejar en el tablero una sola pieza tras cierto número de saltos. Fijémonos en la situación de partida de la ilustración A. ¿Hay alguna manera de garantizar que solamente quede una ficha al acabar el juego? Se ofrece una solución en las ilustraciones B, C y D.

El primer problema consiste en responder a dos preguntas sobre el juego: para garantizar que al final sólo quede una pieza, ¿cuál es el número mínimo de casillas que han de estar vacías al empezar, y en qué lugares de la cuadrícula han de estar? Y si el tablero utilizado fuese de cuatro por cuatro, en lugar de tres por tres, ¿cuántos cuadrados habrían de quedar vacíos, y cuáles habrían de ser?

Examinemos ahora otra versión del juego, esta vez con dos jugadores, a la que llamo Salto-Raptor. Para empezar esta versión, ponemos fichas en todas las casillas del tablero de tres por tres. El primer jugador, llamado Raptor, retira la ficha que prefiera. El segundo jugador, llamado Saltador, hace saltar una de las suyas, si le es posible, y tiene la opción de seguir saltando si puede. Cuando Saltador ha terminado su turno, Raptor trata de realizar sus propios saltos, y así se prosigue por turnos hasta que uno de los jugadores vence al dar un salto que deja solamente una pieza en el tablero. Si, llegado su turno, un jugador se encuentra con una configuración que no permite efectuar ningún salto, arrastra una ficha hasta la casilla central; si tampoco eso es posible, puede arrastrar cualquier ficha no central hasta la casilla vecina que desee.

En las ilustraciones E, F y G podemos ver las tres primeras jugadas de una partida de Salto-Raptor. ¿Quién será el vencedor en este juego, suponiendo que ambos bandos sigan una estrategia óptima?

Solución del problema del mes pasado:

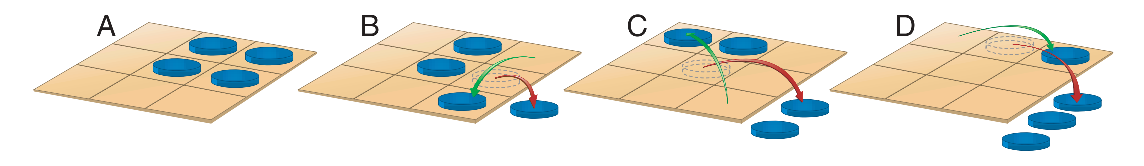
Los problemas de Faroltesta tienen las siguientes soluciones:

1. Juan tiene un rey, y los demás, cartas inferiores.
2. David y Juan tienen reyes, y Carolina una carta más baja.
3. Carolina tiene un rey.
4. Cuando Carolina dice “Yo pierdo” en la segunda ronda, sabemos que Juan tiene una reina y que los otros dos cartas inferiores. David dice “Yo pierdo”, y Juan dice “Yo gano”.
5. Juan tiene un 6, Carolina un 5 y David un 4 o un 5.

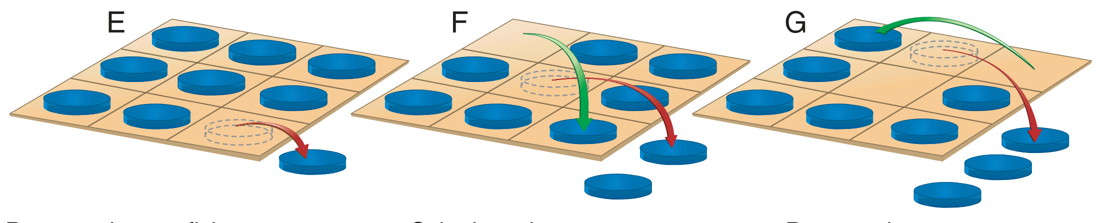
Solución en la Red

Puede verse una solución completa del problema del mes pasado visitando www.investigacionyciencia.es.

VERSION “SOLITARIO”



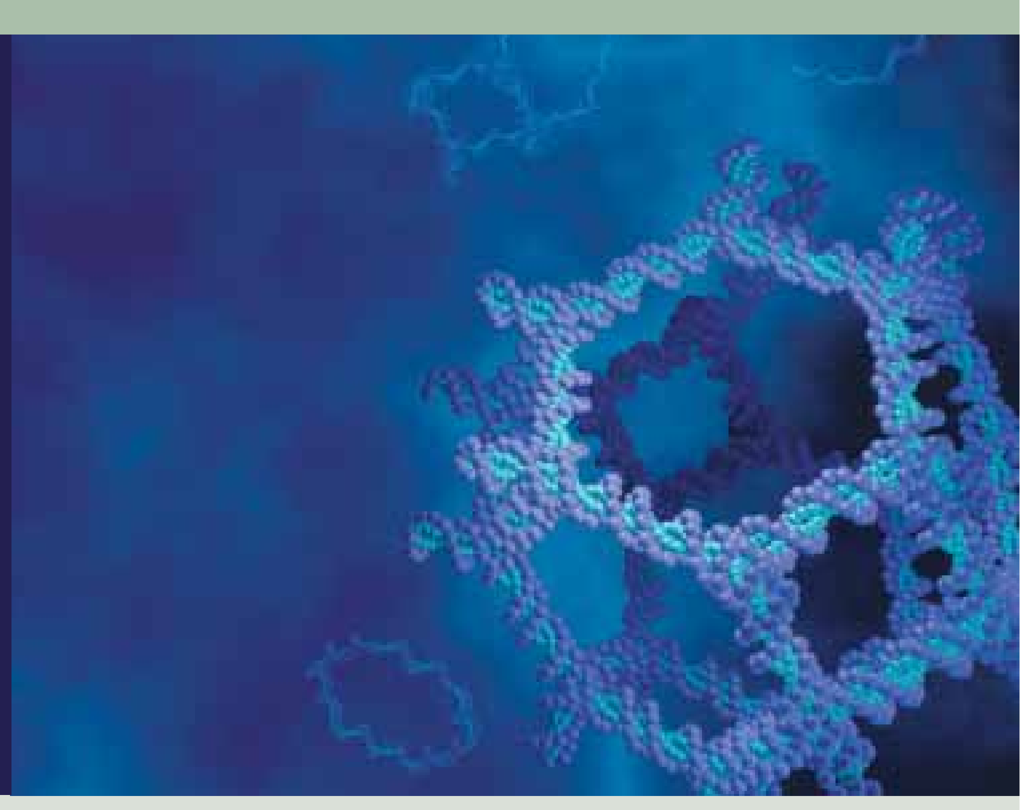
PARTIDA CON DOS JUGADORES



Raptor retira una ficha — Saltador salta una vez — Raptor salta una vez

JOHNNY JOHNSON





LA DOBLE HELICE EN LA NANOTECNIA,
por Nadrian C. Seeman

El ADN no sólo contiene el secreto de la vida. Con él podrían crearse también estructuras y dispositivos de dimensiones nanométricas.



TERAPIAS CON CELULAS MADRE,
por Robert Lanza y Nadia Rosenthal

¿Qué obstáculos se interponen entre las esperanzas que despiertan las terapias basadas en las células madre y los tratamientos clínicos reales?



EXPLOSIONES NUCLEARES ORBITALES,
por Daniel G. Dupont

La proliferación de armas nucleares y misiles hace temer que el sistema mundial de satélites pueda sufrir ataques atómicos.

CONSERVACION DE PLANTAS DE INTERES FORESTAL,
por Antonio Troncoso de Arce, Manuel Cantos Barragán, Juana Liñana Benjumea, Javier Troncoso Mendoza y María García Liñán

Se han ideado y experimentado diversos métodos de propagación y de conservación de especies vegetales del bosque mediterráneo en recesión o peligro de extinción.

SATURNO Y LA SONDA CASSINI,
por Jonathan I. Lunine

Tras siete años de viaje, la sonda Cassini-Huygens por fin escruta de cerca Saturno, sus anillos y su luna gigante Titán.

